

Avoiding traditions that undermine Italy's potential

For years Italy's research council has suffered from scientific direction tainted by political considerations. Changes now in train carry risks that the next generation of Italian scientists will be similarly betrayed.

Italy's national research council, the Consiglio nazionale delle ricerche (CNR), is undergoing one of the deepest reforms that any European research organization has experienced. A new framework law takes away its role as a grant agency, and dissolves its 14 discredited discipline-based advisory committees. So far so good, given that the advisory panels, afflicted by cronyism, had accumulated uncomfortable levels of power which was not always wielded fairly (see *Nature* 394, 712; 1998). But there are ways in which the reforms could still stumble.

The purpose of the changes is to improve the quality of research which, while excellent in pockets, has suffered through poor leadership, which often promoted the mediocre, and weak administration. Research minister Luigi Berlinguer would like to see standards equal to that of neighbouring Germany's Max Planck Society (MPS), and evolution to a system with fewer, but bigger, laboratories of greater scientific stature. To that end, the new system puts more power in the hands of the CNR president (currently Lucio Bianco), whose role had hitherto been weakened by the committees. It creates a seven-member decision-making *consiglio direttivo*, or executive committee, headed by the president and an entirely new scientific committee whose role will be strictly advisory.

Unlike that of the MPS, the executive council is a very political body. Two of its members are selected by the ministry, two by the ministry's own advisory committee, the Assembly of Science and Technology, and two by the CNR president, himself a political appointee. It will have full powers to evaluate institutes, decide on the internal allocation of resources between disciplines and institutes and will have the final say on the appointment of institute directors. It will decide on the new organization of CNR laboratories, whose numbers must be reduced from 300 to 100. It also has the power to set

up new permanent rules governing the establishment, composition and selection of its own scientific advisory structures.

Considering the disastrous effect of political factors on Italy's science over decades, that concentration of power is a cause for deep concern. The role of an independently minded scientific advisory committee becomes all the more important. But the only guideline given by the new law is that half of the scientific committee at the top of the advisory structure must come from within the CNR, half from outside. The worry is that the executive committee might select the scientific committee not just on the basis of scientific qualifications, but also on political criteria. It might turn to the European Commission model, for example, and include representatives of industry and other sections of society. Or it might create a system whereby some scientific committee members are selected by the ministry, as currently, and controversially, happens in France's research council, the CNRS.

Such approaches would be misconceived. The membership of the CNR's scientific advisory committee, and of any subcommittees it sets up, must be restricted to scientists. They should be elected by CNR researchers, just as MPS scientific committee members are elected by MPS scientists. To that end, Bianco would have to withstand resistance from powerful unions which will stand by tradition and demand a vote for all levels of CNR staff, including secretaries and technicians, so diluting the scientific value of the selection.

A fully independent scientific body may go against Italy's grain. But, without such an approach, the CNR will not deserve the financial support it will also require to achieve its goal of outstanding quality. To make the pain of reform worthwhile the CNR will need a budget increase to pay for a much-needed injection of young blood to invigorate its ageing staffing structure. It is, after all, the next generation that must fulfil the purpose of the changes now in progress. □

Primakov's advantageous perspective

Russian science's survival requires the particular attention of the country's new prime minister.

Once again, Russia appears to have stepped back from the brink. Last week's decision by President Boris Yeltsin not to seek a direct confrontation with the lower house of parliament, the Duma, has prompted widespread relief, at home and abroad. Furthermore, his choice of Yevgeny Primakov as the new prime minister, a pragmatist respected for his keen analytical approach to complex problems, offers Russia another chance to put its affairs in order.

Amid his urgent and fundamental problems, Primakov is better placed than most to keep in mind the long term, especially the relevance of his country's dwindling scientific community. The number of working scientists fell from more than 1.1 million in 1989 to 300,000 in 1996. Those who have so far declined either to emigrate or to move into other careers continue to face the indignity of unpaid salaries and vanishing grants. The situation would be even worse if it were not for the

substantial research support from foreign agencies (see page 209). But even some of these are losing faith in Russia's prospects.

Primakov could do much to restore this faith by demonstrating a commitment to a continuing active science base, based on high quality research and linked directly to economic and social priorities. A step in this direction would be to increase pressure for reform on the Russian Academy of Sciences (see *Nature* 387, 533; 1997). Another would be to acknowledge publicly that economic progress requires a healthy system of innovation, which in turn depends heavily on the creative skills of scientists and engineers. Part of the problem facing Russian science is that the political system remains dominated by those with strong links to traditional industries, where interest in research is relatively low. Allowing this sector to continue to control industrial policy would be a recipe for disaster, not only for Russian science but for the rest of the country too. □



Brussels seeks BSE diagnostic test to screen European cattle

[PARIS] The European Commission is to spend about ECU 1 million (US\$1.16 million) on organizing a competition to select a diagnostic test for bovine spongiform encephalopathy (BSE) in cattle. This is a preliminary step towards the ultimate goal of being able to screen each of the millions of animals entering Europe's abattoirs.

The move is part of a panoply of measures being considered by Brussels to meet the threat that, although draconian measures in the United Kingdom promise to end the BSE epidemic there, the lack of precautions in several other European countries may be providing a breeding ground for the disease.

For example, although the United Kingdom and France have introduced a blanket ban on the feeding of meat and bone meal to all farm animals, the European Union (EU) has only succeeded in introducing a ban on the feeding of ruminant proteins — in meat and bone meal — to ruminants. Feed must also be heated at 133 °C for 20 minutes.

The United Kingdom, France and Belgium have also implemented a ban on sheep heads, spinal cords and spleen in the human food chain, but an EU-wide ban has been opposed by Germany and other countries on the grounds that they do not have BSE. Franz Fischler, the EU agriculture commissioner, is likely to resubmit the proposal for such a ban to Europe's farm ministers soon. He will be backed by a report on the risk of BSE in sheep that is scheduled to be adopted this month by the commission's Scientific Steering Committee, the body that oversees all the commission's scientific advisory committees (see *Nature* 395, 6; 1998).

"The risk of human exposure to BSE is probably higher in those countries which are officially BSE-free, and subsequently take no precautions," says Gérard Pascal, chairman of the committee. He adds that cattle still pose a risk in certain European countries.

Marc Savey, a researcher at the French National Veterinary Centre, points out that the occurrence of sporadic cases of BSE in cattle, despite the ruminant feed ban, shows clearly the existence of cross-contamination with chicken and pig feed, or fraudulent use of such feeds for ruminants. He adds that a blanket ban on the feeding of meat and bone meal to all farm animals is needed to ensure that this source of infection is cut off.

It is in this context that the commission has also launched a programme to develop tests capable of identifying the presence of mammalian proteins in composite feed supposed to be free of such proteins, and of checking that feed has been heated correctly.

The risk is not that of new BSE epidemics on the scale of millions of cases as occurred in the United Kingdom, but that low levels of BSE may persist in European herds, says Pascal. Even with a few hundred cases a year, there is a risk of recycling the disease through feed, and of allowing diseased carcasses to enter the human food chain, says Savey.

"It is impossible that countries such as Germany, Italy and Spain have not had cases," asserts one of the commission's advisers on BSE. Given the flow of contaminated feed and animals in these countries (see *Nature* 382, 4; 1996), "to believe that BSE suddenly stops at the Franco-German border is scientifically ridiculous".

To counter the threat that BSE might persist in continental Europe, the commission is now giving a "high priority" to a diagnostic test, according to one official. "The commission's aim is to have a test applicable to every animal arriving at the abattoir," says Pascal, adding that the test must detect animals that have BSE but do not show clinical symptoms.

In a related move, the UK government's Spongiform Encephalopathy Advisory Committee last week recommended surveying more than 1,000 cattle brains from slaughtered animals to see if they are harbouring BSE without showing symptoms, using a test developed by John Collinge at Imperial College School of Medicine in London.

The European Commission's competi-

tion for a diagnostic test was launched in July by its consumer directorate (DG24). Each test must have a solid scientific basis, be at an advanced stage of development, and be able to be scaled up quickly for industrial use with millions of cattle, according to an official. The commission expects the competition to be completed by the end of the year.

Under the blind-study protocol, the winners will be sent thousands of coded samples of cattle brain and other tissues, but will not be told which are infected. Four out of 10 applicants have now been selected.

Commission officials declined to identify the winners, but the Zurich-based biotechnology company, Prionics, this week confirmed that it has been selected. A pilot study of the Prionics test (see *Nature* 390, 74; 1997), in screening 3,000 cattle from Swiss slaughterhouses, was approved by Swiss veterinary authorities last month, and the results are expected at the end of this month.

Another likely candidate is the Irish company Enfer Scientific which has licensed technology from the UK company Proteus for the development of a post-mortem test of BSE in cattle. The Proteus technology is also being studied by the UK Public Health Laboratory Service, as a potential diagnostic test for Creutzfeldt-Jakob disease in humans.

Current tests for BSE are based on the use of antibodies to detect the presence of the abnormal prion proteins that accumulate in

Don't copy our clone, say Dolly scientists

[LONDON] Pictorial — as well as physical — copies of Dolly the cloned sheep (right with 'foster mother') will in future need special permission following a decision by the Roslin Institute in Scotland to apply for trademark protection to control the use of her picture in advertising.

The decision was made after Zanussi, the electrical appliances manufacturer, used a picture of a sheep in an advertisement with the title: "The misapppliance of science".

But the advertisement may not have had the desired impact. Surprisingly, almost half of those responding to a UK-wide survey whose results were released this week said they had not heard of Dolly.

But, when given a series of six possible options, two-thirds of the survey respondents correctly answered that Dolly was the first mammal cloned from an adult cell. Some 49 per cent of respondents thought that Dolly had been cloned to "advance human cloning", a goal approved



of by only 11 per cent of those questioned.

Although most respondents approved of biotechnology to develop medicines, only 17 per cent approved of genetically modified food. The survey was carried out by the public relations consultancy HCC-De Facto Group.

Ehsan Masood

infected tissues and cause the clinical symptoms of spongiform encephalopathies. Bruno Oesch, from Prionics, claims its test is "99.9 per cent specific", and says that the company has reduced the time from processing samples to obtaining results to six hours.

But several scientists are sceptical of the performance of all the tests being developed, arguing that there remains uncertainty about their specificity and sensitivity.

Diagnosing spongiform encephalopathies is intrinsically difficult, compared with an AIDS test for example, where the presence of antibodies to HIV clearly indicates the presence of the virus. Abnormal prions may only be present in animals at a relatively late stage, so detecting their absence may not guarantee that the animal is not sick, argue several scientists.

They claim that, whereas tests may be able to detect BSE in sick animals, so far no published data provide convincing proof that tests could reliably detect preclinical cases. The commission's ambition of testing millions of cattle is "very premature", says one.

Despite these limitations, the potential of a test as a tool for managing BSE is "so great that the commission has put a very high priority on it," according to one official from the commission's consumer directorate.

To put pressure on BSE-free countries to accept stricter precautions, the commission has set up an expert panel to prepare a "geographical risk assessment" of BSE in Europe. This will construct mathematical models of risk based, for example, on feed use, imports and farming practices. Using such a model, Germany, currently listed as BSE-free, would probably be reclassified as 'at risk' from BSE, says one commission official. **Declan Butler**



Fischler: weighing BSE risk in sheep.

Cuts 'could leave energy department brain-dead'

[WASHINGTON] The US Department of Energy (DOE) is facing a crisis in overseeing its research programmes as its most experienced technical staff leave and others await retirement, according to senior officials.

Almost half of the DOE's 500 technical managers — who oversee billions of dollars worth of research programmes from its Washington headquarters — will be eligible for retirement in the next five years, according to a report presented last week to the department's Laboratory Operations Board.

"The technical managers are overwhelmed and discouraged about the future," says David Crandall, manager of the inertial confinement fusion programme and co-chair of the panel that produced the report.

"Due to continued downsizing, without concomitant reduction in programmes, those who remain are overwhelmed with work and often exhausted," the report says. It also warns that the "pipeline" of younger people to become research managers at the department "is virtually empty".

The energy department has been under heavy pressure to reduce its staffing levels since 1995, when the then newly elected Republican Congress threatened to abolish it outright. Hazel O'Leary, then energy secretary, responded to this threat by proposing a Strategic Alignment Initiative (SAI), which committed all sections of the department to meet fixed targets for staff reductions.

As a result, the department is unable to replace people who leave, including the technical managers who are supposed to supervise its physics, nuclear weapons and other energy-related research programmes.

"The department has a potential crisis within five years, in which time it will go

brain-dead," asserts Richard Hopf, a deputy assistant secretary for procurement.

At the meeting of the Laboratory Operations Board in Washington on 9 September, other senior officials said they were already aware of the problems raised by the report.

Victor Reis, for example, assistant secretary for defence programmes, said the nuclear weapons stewardship programme has a dozen technical management positions that have gone unfilled for a year because of the SAI targets. "Everybody just took a proportional cut" in staff numbers, Reis said. "It wasn't related to the job that had to be done."

John McTague, co-chairman of the Laboratory Operations Board, who is a vice-president of Ford and once served as science adviser to George Bush when he was US president, pointed out that the issue of attracting good people to work for the federal government was a perennial one.

But Martha Krebs, assistant secretary for energy research, echoed Reis's frustration with the current logjam.

According to Crandall's report, the problem of retiring staff is most acute in Krebs's office, which oversees \$2 billion worth of research in physics and basic science. Of 102 technical managers at the office in 1995, 14 have already left and another 33 will be eligible to retire by the end of this year. Krebs says the root of the problem lies with Congress, where appropriations committees have demanded staff cuts and tried to block efforts to second technical staff from the laboratories to the department's headquarters.

But staff at the committees claim that although the department was invited to propose a plan that would enable it to fill key positions, it has failed to do so. **Colin Macilwain**

Biology teaching wins \$90m boost from Howard Hughes institute

[WASHINGTON] The Howard Hughes Medical Institute (HHMI) has awarded \$90 million in grants to support the development of undergraduate education in biology at 58 US universities.

The money — the largest-ever tranche of education funding from the institute — is intended to help the universities to introduce undergraduates to research, to update undergraduate science courses, and to build outreach programmes with local schools and community colleges.

Joseph Perpitch, HHMI's vice-president for grants, says the money will encourage improvements to undergraduate education at a time when major research universities have been criticized for neglecting it.

A commission established by Ernest Boyer, the late president of the Carnegie Foundation, reported earlier this year that many leading research universities were failing to teach undergraduates what research was, or to let them do any of it (see *Nature* 392, 746; 1998).

Perpitch says that the 58 universities selected by HHMI for support from 191 applicants "have a strong commitment to undergraduate science education". He adds that the Boyer report "didn't reflect what's happening at these institutions".

At Cornell University in New York State, for example, a \$2.2 million grant will support undergraduates doing laboratory research, and help to continue an outreach

programme to high school biology teachers.

Peter Bruns, director of the Cornell programme, says undergraduates will be exposed to laboratory practice on several levels as a result of the grant. All 1,000 students on an introductory course in biology will spend a short time in a research lab. Three hundred of the students will also have the chance to conduct independent research, and 70 will receive a stipend to spend a summer with a mentor in the lab.

But Bruns says the overall picture for undergraduate education remains "mixed" at Cornell and other research universities. He thinks many institutions "have not come to grips with the issue" of undergraduate science education. **C.M.**

Human genome deadline cut by two years

[WASHINGTON] The main US government sponsor of the Human Genome Project adopted an ambitious plan this week to finish sequencing the human genome by 2003, two years ahead of the original target.

The advisory council of the National Human Genome Research Institute (NHGRI) unanimously approved the plan, which sets as an interim goal the production by the end of 2001 of a 'working draft' of the genome that is at least 90 per cent complete (see *Nature* 393, 399; 1998.)

The plan also calls for one-third of the genome to be completely sequenced by 2001. It calls for a focus on gene-rich regions as this interim goal is reached, and suggests that an international peer review process be set up immediately to prioritize regions to be sequenced, based on the needs of the international research community.

The plan, which sets goals for the entire federal effort between 1999 and 2003, is to be published in full next month. As recently as May, a draft version contained no reference to advancing the target date to 2003.

"We are talking about big and ambitious, even audacious, plans," said Francis Collins, the NHGRI director, at a meeting of the advisory council on Monday (14 September). He described the new targets as "a stretch". But he said the accelerated timetable was vital because every year that passes without the genome sequenced was "an opportunity lost for biomedical research across the board". Collins added: "This is not a time to be conservative, cautious or to coast along."

The new plan will require government-sponsored sequencers to double or even triple their output. But Collins said that it does not rely on the institute receiving additional funding from Congress beyond regular annual increases "in the neighbourhood of 10 per cent".

To accommodate the new goals, he said, the proportion of the institute's \$218 million budget spent on sequencing will grow from its current 45 per cent to 60 or 65 per cent by 2001.

Collins denied that the plan was formulated in reaction to the announcement in May by Celera Genomics, a company headed by genome sequencer J. Craig Venter, that it aims to sequence the genome within three years (see *Nature* 393, 101; 1998). "This is action, not reaction," said Collins. But he added that the goal of producing a 'working draft' presents "great potential for collaboration" with Celera.

Venter agrees. He calls the government's plan for the draft "a tremendously positive development". It "certainly alleviates any critics' questions or problems about [Celera's] ability to assemble the whole genome".

Celera's plan has been questioned

because it relies on the technique of whole genome 'shotgun' sequencing that many predict will present significant reassembly problems. The rapid production of a 'working draft' by the government would provide mapping information of great value to the private reassembly efforts.

David Bentley, head of human genetics at the Sanger Centre in Cambridge, United Kingdom, which will compile one-third of the draft sequence due by 2001, warmly endorsed the new plan, and called the international peer review proposal "excellent". He

added: "There is a need to have some sort of process of reviewing all these claims to ensure that the right regions are done, and the best use is made of the resources."

Collins stressed that the 'working draft' was not a "throwaway" product, but an intermediate step to the completion of the genome. "There should be no additional expense created" by its generation, he said.

The Department of Energy has approved the five-year plan. The department, a joint author of the plan, is the other federal sponsor of the genome project. **Meredith Wadman**

Snowball comet idea refuses to melt away

[WASHINGTON] Twelve years after two space physicists proposed that Earth is being constantly bombarded by small, watery comets, most scientists are ready to abandon the search for such objects. But at least one observation is still planned to try to confirm or refute their existence.

Louis Frank and John Sigwarth of the University of Iowa, whose controversial claim has been attacked repeatedly in journals and at scientific meetings, still believe house-sized 'snowballs' are entering the atmosphere every few seconds.

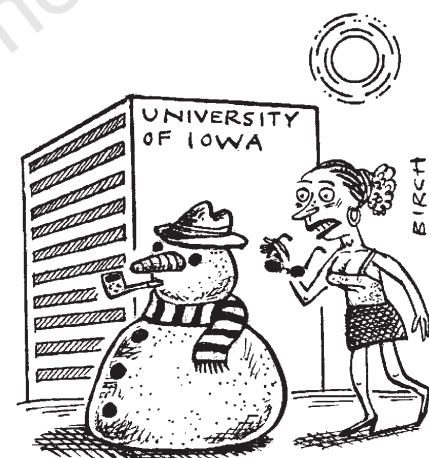
Their theory — based on dark spots that showed up in ultraviolet images of Earth taken by the Dynamics Explorer satellite in the early 1980s — appeared vindicated when a second spacecraft, Polar, saw similar spots after its launch in 1996.

Former critics expressed cautious support for the idea, and the director of the US space agency NASA's Goddard Space Flight Center set up a small, informal steering group to help orchestrate attempts to prove or disprove the theory.

Since then, however, a steady stream of evidence has gone against the small comet interpretation. The University of Arizona's Spacewatch asteroid-searching telescope, which should have revealed about 2,000 objects in seven years of scanning the skies, has turned up nothing. Other scientists measuring water vapour in the stratosphere say there is not enough to support the claimed frequency of comet bombardment.

The latest and perhaps most damning refutation is published next month in *Geophysical Research Letters*. Based on their analysis of a day's worth of raw data supplied by Frank, Forrest Mozer and James McFadden of the University of California at Berkeley conclude that the spots in the Polar images were produced by the instrument and the way in which the data were processed. Frank rejects that interpretation.

But more bad news is on the way. Robert Meier of the US Naval Research Laboratory



has nearly completed his analysis of data collected by a navy space radar network last autumn. In one six-week period, he says, the radar should have seen tens of thousands of objects. Instead it saw nothing.

Robert Hoffman, the project scientist for Polar and leader of Goddard's small comets steering group, says at least two pieces of evidence have yet to come in. The Spacewatch team has taken 17 images in an observing mode specially designed to capture the objects, if they exist, but has yet to release its analysis.

Hoffman hopes to arrange an independent review of those images, which Frank says should prove his theory. Another team led by John Mathews of Pennsylvania State University, with support from NASA and the National Science Foundation, plans to use the Jicamarca Radio Observatory in Peru to hunt for the objects.

Beyond that, however, few scientists seem interested in pursuing the matter further, and Hoffman is not even sure his steering group will meet again. Although he says there has been no "knockout punch" for the small comets theory, his own group will not attempt a "verdict" one way or the other. "I think the burden of proof is on Frank right now," he says. **Tony Reichardt**

German poll could change research goals

[MUNICH] With Germany's federal election looming, parliament has approved a 3.4 per cent, DM500 million (\$295 million) increase in next year's education and research budget, as proposed by minister Jürgen Rüttgers.

But the public seems unimpressed. Polls indicate that the current Christian Democrat/Free Democrat coalition government may not be running the country after the elections, due on 27 September.

A late turnaround remains possible, but most money at the moment is on either a Social Democrat/Green coalition or a 'grand coalition' of Christian Democrats (CDU) and Social Democrats (SPD).

For scientists a grand coalition would mean little change, as both major parties essentially share a platform on research. Both pledge at least to maintain current levels of funding for basic research organizations such as the Max Planck Society and the Deutsche Forschungsgemeinschaft (DFG), Germany's main grant-giving body. Both are committed to competition and industrial relevance in research. Both say genetic and information technologies are priorities, and agree that medium-sized enterprises should be encouraged to conduct more research.

One main difference is that the SPD wants to double the budget for science and education in the next five years. But most scientists take this promise with a pinch of salt. "There certainly is a need for a change, but I doubt that the SPD would keep its financial promises," says Siegfried Bethke, head of particle physics at the University of Aachen.

Rüttgers promised to increase the science budget when his appointment as 'minister of the future' was announced in 1994 amid great fanfare. In fact, the budget remains at

much same level and has fallen around five per cent in real terms.

The SPD wants to move responsibility for research from the education ministry to the economics ministry, which would be run by CompuNet founder Jost Stollmann, 43, a businessman with no political experience.

The research portfolio would be given to Edelgard Bulmahn, a 47-year old political sciences graduate with 12 years' experience on the parliamentary committee for education and research. Bulmahn says that one of her goals is to work closely with other ministries. "The SPD wants research policy to be fully integrated with other government policies, relating for example to tax, environment and transport," she says.

Prospects of a coalition including the Green Party sends a wave of fear through the scientific community. Although the Greens have softened their stance on biotechnological research (see *Nature* 392, 213; 1998), their election programme states that the party "generally rejects genetic engineering".



Bulmahn: SPD wants research integrated with policies on tax, environment and transport.

The Greens want to tighten Germany's genetic-engineering laws and ban field trials of genetically modified crops. Manuel Kiper, spokesman for the Green's parliamentary group on research policy, calls field trials "a game of hazard with evolution". Both policies would cause conflict with their potential coalition partner, as the SPD has a much less hostile attitude to genetic engineering.

Kiper says that, although the Greens would respect existing treaties, they would want to start negotiations immediately to stop Germany's involvement in projects like the international fusion project ITER, the International Space Station (ISS) and, of course, the nuclear energy programme.

The SPD is broadly sympathetic, but hopes that international support for ITER and ISS will fade with time and calls for Germany's 19 nuclear power stations laws to be shut down gradually over the next 30 years.

The Greens want to promote environmental research, alternative energy research and complementary medicine. They also want to bring high-profile private environment research institutes — such as the Eco Institute and the Institute for Social Ecology — into the state funding system, with government providing 25 per cent of the institutional funding. The SPD supports this plan.

As a junior coalition partner, the Greens would be allocated at least two ministries. Their first priority would be to win the environment ministry. One of their second choices would be research, although the SPD is against this idea.

Ernst-Ludwig Winnacker, president of the DFG and long-time lobbyist for genetic engineering in Germany, says that scientists are "quite worried about the possibility of a coalition with the Greens". Having the Greens in government would be quite detrimental to research, he says.

Not all scientists share his opinion, however. Bethke, for example, says he has "a very good feeling" about the Greens' support for basic physics research, such as the planned new linear accelerator in Hamburg.

Whoever takes over research and education after the election will face two major unresolved problems, which no party has addressed in its election campaign. The first is the chronic underspending on university building and large equipment. According to the Wissenschaftsrat, Germany's science council, the annual budget for this needs to be increased by at least DM1 billion to meet what it regards as urgent needs.

The second problem is east Germany. Although more than DM20 billion has been invested in its research infrastructure since 1990, it still lags behind the west, which still has four times more researchers per head of population. **Quirin Schiermeier & Alison Abbott**

Plan to pool part of Europe's science cash

[MUNICH] The heads of Europe's national research councils, who meet regularly as a group called the Eurohorcs, are to consider a proposal to pool part of their funds for reallocation to European-level programmes.

The idea has been put forward by Reinder van Duinen, president of the Dutch research organization NWO. It is aimed at promoting research in areas not supported through the European Commission in Brussels, and would probably be run by the European Science Foundation (ESF). Specific ideas include phase 1 clinical trials and some areas of social science.

Van Duinen says he is "encouraged" by the response he has received to his proposal, which emphasizes the importance of a light, non-bureaucratic structure.

Ernst Winnacker, president of the Deutsche Forschungsgemeinschaft (DFG), is

enthusiastic. But he warns that the legal implications of transferring money allocated for national research to an international body could be a stumbling block in some countries.

Richard Brook, head of both Eurohorcs and the UK Engineering and Physical Sciences Research Council, is cooler. "There are already mechanisms for setting up international collaborations between national research organizations through bi- and multilateral agreement," he says.

Van Duinen says such collaborations take place in an ad hoc way — "the wheel is constantly being reinvented". He adds: "We need a system which has a memory of the basic organization of collaborations."

The Eurohorcs will discuss the proposal at their next meeting, to be held in Oslo in October.

A. A.

Foreign contracts 'Russia's only chance'

[MOSCOW & LONDON] The appointment of Yevgeny Primakov as the new Russian prime minister and the promise to give cash-starved scientists a portion of their unpaid salaries may go some way towards stabilizing Russian politics. But they are unlikely to ease the hardship for researchers in the foreseeable future.

No solution is in sight to the salary crisis among scientists (see *Nature* 395, 109; 1998). The Ministry of Atomic Energy (Minatom), for example, is owed 2.8 billion rubles (US\$280 million) by the government, including 800 million rubles in up to four months of unpaid salaries. Staff staged a mass picket outside Minatom's Moscow building earlier this month.

Victor Ivanov, Minatom's newly appointed deputy minister, acknowledged last week to a press conference in London that foreign contracts are the only realistic solution to the crisis. Ivanov said that the ministry and its local reactor management had dug deep to support those most in need, but there is little more that can be done.

Overseas orders for nuclear materials, such as radioactive isotopes, "are the only way our institutions will survive", he said. Speaking on the day Primakov was nominated as prime minister by President Boris Yeltsin, Ivanov said Minatom institutions that relied solely on the Russian government "will have to shut down, or suspend work".

But Ivanov stressed that Russia was to press ahead with plans to expand its domes-



Unlucky strike: nuclear industry workers picket the parliament building in Moscow last week in support of demands for four months' back pay. Slogan on the left reads: "A hungry nuclear worker is much more dangerous than any other one".

ALEXANDER ZEMLIANICHENKO/AP

tic nuclear power programme. Three pressurized water reactors are being completed, and regulatory as well as local planning approval had been obtained for a fourth. He also said that Russia was keen to develop a new generation of fast breeder reactor with the help of foreign partners.

Ivan Gradobitov, deputy chairman of the nuclear workers' trade union, said that the government could not shirk its responsibilities. "The ministry is responsible for the situation since it serves as a state customer

for its own enterprises," he said.

Meanwhile, the Russian Committee of Scientific Collectives (RCSC), a federation of science trades unions, continues to send telegrams alerting Yeltsin to the catastrophic situation in Russian science, and repeatedly demands that the cabinet honour its earlier promise to pay scientists' salaries in full. But its exhortations fall on deaf ears.

The unions are also calling for the reinstatement of Vladimir Fortov, who was sacked as minister of science and technology in April, although Fortov is understood to have ruled out a return to office. Izosim Molchanov, the sacked deputy minister responsible for science within the finance ministry, has also yet to be replaced.

At the last RCSC meeting, union members voted for a series of strikes outside government buildings in Moscow. Unless the government acts by the end of September, staff from research institutes surrounding Moscow will blockade three highways leading to the city. The strikes will culminate in a series of protests across Russia on 7 October.

The Russian cabinet's decision of 17 August to freeze all short-term payments has crippled the country's banking system and left citizens with effectively no access to their money.

Bank accounts have been frozen, partly because the banks are running short of money, but partly too as collateral against government debts, and partly because they have converted rubles into dollars to safeguard the value of their deposits. The closure of accounts means bills cannot be paid, depriving the government of its usual sources of revenue.

Shops are emptying quickly as those with cash rush to stock up on food before the next price rise.

Carl Levitin & Ehsan Masood

Hopes ride on manager with a scientist's mind

[MOSCOW] The hopes of an ailing Russia are riding on Yevgeny Primakov, the country's so-far scandal-free prime minister, who has combined his career in politics with success in several other fields.

A former economist, journalist, Middle East expert, head of two research institutes, spymaster and diplomat, Primakov caps his impressive range of expertise with Russia's highest scientific rank: that of academician.

Between 1962 and 1970 he worked as a journalist specializing in the Middle East, on which he later wrote his doctoral thesis. He was elected a full member of the Russian Academy of Sciences in 1989, the year he became head of the KGB.

He is well known politically for his analytic approach to problem solving. It is a skill that will be severely tested as he charts a course through Russia's financial crisis and attempts to reinstate – and, even harder, maintain – a regular flow of salaries to scientists.

Yuri Emel'yanov, a former senior member of Russia's scientific civil service, thinks Primakov will succeed where others have failed. Emel'yanov remembers Primakov's handling of conflicts while at the helm of two of the Russian Academy of Sciences' research organizations – the East Research Institute and the World Economy Research Institute – from 1970 to 1989.

"I witnessed how he solved some of the most

delicate problems arising in the scientific community," says Emel'yanov. "Each time he acted as any scientist should: analysing a problem in depth and trying to understand it in its broader context."

According to Vitaly Goldansky, a prominent Russian physicist, Primakov tends to draw on the advice of expert advisory committees, and will use his contacts from the world of research to gather around him some of Russia's finest experts.

"He is a born manager," says one of his subordinates. "It makes no difference for him what kind of organization he is running – a scientific institute, a foreign ministry, an intelligence service or a local hospital."

C.L.

US scientists call for fairer animal funding

[WASHINGTON] Scientific societies in the United States are challenging a government policy that they claim is harming biomedical research by selectively punishing investigators who work with animal models.

Last week, 75 societies, led by the American Association of Immunologists and the American Physiological Society, wrote to Harold Varmus, director of the National Institutes of Health (NIH), and William Raub, a top science policy adviser in the Department of Health and Human Services, to protest at guidelines on repaying overhead costs of animal research facilities (ARFs).

The letter attacks the current policy, under which these may not be billed to an institution's general overheads. Many institutes have responded by billing investigators directly. The societies call this situation "a hindrance to progress" on diseases that require animal models, because it makes proposals using animals seem more expensive.

The problem arises from the fact that the government, led by the White House Office of Management and Budget (OMB), has in recent years set about enforcing — albeit very unevenly — a policy that has officially been on the books since 1979.

Institutions including Harvard University and the University of Chicago have been forced to bill investigators directly for overhead costs at their ARFs, whereas some, such

as Columbia University, remain unscathed. Others, such as Yale University, have had to stop charging overhead costs to the indirect cost pool, but have cushioned their scientists by covering the charges from other university funds.

Scientists believe that this has created an uneven playing field for proposals going before NIH study sections, where reviewers may be tempted to select a grant applicant who can complete a project at a lower cost.

The government's accountants maintain that ARFs are "specialized service facilities", which, like wind tunnels and research reactors, cannot charge overheads to the indirect cost pool. The system, the government believes, not only makes scientists using the facilities more cost-conscious but also prevents the broader community subsidizing expensive, highly specialized facilities.

But scientists say that classifying ARFs as "specialized" no longer makes sense, as advances in animal research in the past two decades — in particular the explosion in work on transgenic and knockout mice — have changed ARFs from holding stations for animals into genuine research space, just like any investigator's laboratory.

"The research community has really been almost speaking with one voice on this," says Judith Vaitukaitis, the director of the National Center for Research Resources (NCRR) at

NIH. "What they are saying is that research with animals is like research with test tubes, or any other part of biomedical research, and should be treated as such."

Encouraged by a recent report from the National Research Council (NRC) endorsing its views, the scientific community is taking on the government. "Scientists will become reluctant to pursue important lines of research involving animal models out of concern that their research has less chance to be funded," says its letter. It urges the government to reverse course and adopt the recommendations of the NRC report, which was requested by NCRR and released in June.

The report calls the current policy "misguided" and urges the government to allow institutions to meet ARF overhead costs from the indirect cost pool. The authors see "no reason to single out animal research". They estimate that implementing the policy doubles the average animal housing rates charged to scientists.

Varmus, Raub and cost-accounting officials from the Department of Health and Human Services were due to be briefed on the report by its authors on Monday (14 September). Raub said before the briefing that the report would receive a "very serious look" from the department's officials. But any change in policy would have to win the approval of OMB.

Meredith Wadman

Spain lets Latin America 'repay' debts by protecting environment

[BARCELONA] Spain has agreed to a US\$70 million 'debt for nature' swap to aid biodiversity conservation in Latin America, in what government officials call "the most ambitious project in the history of Spanish international cooperation".

Latin American countries that owe substantial sums to Spain will be allowed to use the money to protect and improve the local environment. The project is called Araucaria, after an Andean conifer, and the money will support the development of communities living near some of Latin America's most important ecosystems.

All improvements carried out with funds owed to Spain will be performed with Spanish technology and equipment.

The five-year project will be officially launched next month. Prince Felipe gave it his support during a visit last month to the Galápagos archipelago, where Spanish Foreign Office technicians and scientists are drawing up schemes for inclusion.

This is the first time an economic mechanism has been set up in Spain to cancel a substantial part of Latin American countries' debt in return for a commitment to preserve the natural environment.



City beggars in Venezuela: the new scheme could help people earn a living in their home villages.

One of the main purposes of Araucaria is to protect biodiversity in areas especially sensitive to human action. Researchers are trying to work out how to promote the development of local communities without putting biodiversity in jeopardy.

"It is a pioneer initiative that responds to a historical plea of Latin American debtor countries," says Jesús García, head of the Institute for Latin American Cooperation. "Spain is fulfilling a commitment it made in 1992 during the Earth Summit conference in Rio de Janeiro."

A typical Araucaria project is the construction of a road joining Iquitos, the

capital of Peruvian Amazonia, and Nauta village, which stands at the entrance to the Pacaya-Samiria nature reserve. This will open up non-destructive methods of production by increasing tenant farmers' access to previously inaccessible areas.

José María Aznar, Spain's prime minister, will visit Iquitos this month to inspect Araucaria's activities in the area. Other projects are under way in the Colca valley in Peru, and in Savegre valley, Costa Rica.

"Araucaria will lead to concrete links between economic issues and those concerning the conservation of the natural environment," says Alvaro Rengifo, head of the Department of Foreign Investments. "It will allow the experience of projects already running in Costa Rica and Peru to be reproduced in other countries."

Many experts see such agreements as a strategy that can help preserve one of the Earth's most important regions of biodiversity. The task will not be simple — it will require complex legal steps, as well as the approval of the parliaments involved.

García says the main advantage of Araucaria "is its coordinated and selective nature".

Xavier Bosch

JOSE CARUCI/AP



Excellence rewarded: NSF's Rita Colwell hands Winner E. Alexander an award for mentoring.

Clinton seeks refuge in calls for mentors to help minorities

[WASHINGTON] The beleaguered US president, Bill Clinton, has suggested that every scientist on a federal grant should be a mentor to guide minorities, women and disabled people into careers in science and technology.

At a ceremony at the White House last week to present the 1998 Presidential Awards for Excellence in Science Mathematics and Engineering Mentoring, Clinton said he was asking the National Science and Technology Council to look at ways "to achieve diversity throughout our scientific and technical workforce" and to report back to him within six months.

Clinton, who had just met with the senior Democrat senators whose support he needs to save a presidency now gravely imperilled by the Monica Lewinsky scandal, told the awardees that he "couldn't think of a better subject" than scientific mentoring for him to address "during this process that I'm going through".

Clinton said there were "tens of thousands of potential mentors" among university and laboratory scientists who are supported by the federal government. "If every one of them was to become a committed and dedicated mentor, think what that would mean," he said.

Arguing for more action to attract women and minorities into science, Clinton cited a report from the American Association for the Advancement of Science, which finds that the number of African Americans and Hispanics entering graduate science education fell by 20 per cent between 1996 and 1997.

The association says court rulings and moves by the University of California to end affirmative action have led to "an unwelcome environment" for minority graduate students, causing the decline. **Colin Macilwain**

Japanese institutes wary of increased autonomy

[TOKYO] Japanese scientists are expressing concern about the government's decision to carry out a large-scale reorganization of national research institutes attached to science-related ministries. The move will include transforming more than half of the institutes into semi-autonomous 'agencies' with administrative independence from the government (see *Nature* 389, 897; 1997).

The government says the change is needed to turn the institutes into genuine centres of excellence. But in late August an association representing the directors of Japan's 93 national research institutes released a report arguing that research institutes are fundamentally different from other government organizations, and that certain management procedures — especially those with strong emphasis on performance-related targets — would be inappropriate for them.

The plan is part of the government's drive to improve the country's administration. The government says its plans to turn the institutes into semi-autonomous bodies is based on the 'Next Steps' strategy developed in the United Kingdom. Each institute will have a separate management system and an external assessment body made up of representatives from the private sector to evaluate its performance.

The administrative reform plan has taken a sluggish path since its approval in December. But it has advanced rapidly in the past month in the wake of the appointment of the new prime minister, Keizo Obuchi, who has pledged to start implementing the reforms as soon as possible.

Although individual ministries and agencies have started to draw up measures for reorganizing the national institutes, many are concerned that the views of scientists working in the institutes will not be given much attention.

Under the plan, 54 of the 93 national research institutes will be reorganized. These include the Science and Technology Agency's (STA)'s National Institute of Radiological Sciences, the Environment Agency's National Institute for Environment Studies, and all 15 institutes of the Ministry of International Trade and Industry's Agency of Industrial Science and Technology (AIST).

Institutes carrying out similar research are likely to be merged, and those with poor track records will be closed down before new agencies are formed.

But directors of national institutes have repeatedly voiced their concern about the plan. In particular, they point out that many of the institutes carry out basic research and are therefore unlikely to yield profitable

returns. As a result, they argue, a management approach that emphasizes financial performance-related targets would become a burden on researchers.

"The government's effort to promote scientific research and its effort to rationalize the institutes seem to be conflicting with one another," says Shinichi Ohashi, director of AIST's National Institute of Biosciences and Human Technology.

The new status of the institutes would give them more flexibility in funding. For example, they would be able to carry over unspent research funds to the following fiscal year, in contrast to the current rules, under which annual funds must be used up within the year. But scientists are concerned that the change would make the institutes more vulnerable to cuts in government funding.

Michinori Nishio, chairman of the Association of Directors of National Research Institutes and director of the agriculture ministry's National Institute of Agro-Environmental Sciences, warns that, once institutes lose their direct links with government, they would also be in danger of losing a secure source of funding and have to start relying on other sources.

The association's report asks for continued funding for research institutes after their launch as agencies, and the Science and Technology Council (STC), Japan's main science policy-making body has urged the government to ensure that the new system does not affect the distribution of government funding to the institutes, even if they are free to obtain funding from external sources.

Despite government confidence that the reform will help the institutes to run more efficiently, many doubt that the new agencies will be any different from existing semi-governmental organizations (*tokushu hojin*), such as STA's National Space Development Agency and the former Power Reactor and Nuclear Fuel Development Corporation.

The *tokushu hojin*, which are funded by both the government and private sectors, have often been criticized for their poor management, lack of transparency, and inefficient use of funds. The Institute of Physical and Chemical Research (RIKEN) is the one exception that is widely seen to have made a success of the greater autonomy conferred by *tokushu hojin* status.

Previous attempts to reform these organizations have failed (see *Nature* 373, 551; 1995). STC's request states that the government would first have to reorganize the nation's semi-governmental organizations if they were to create real centres of excellence in Japan. **Asako Saegusa**

Consumers given louder voice on food safety in Britain

[LONDON] Britain's Ministry of Agriculture, Fisheries and Food (MAFF) has completed the appointment of consumer representatives to all its advisory committees involved in food safety, with the nomination of four lay members to its Advisory Committee on Pesticides and sub-groups. According to Jeff Rooker, the food safety minister, "these appointments mean we can reassure consumers their interests are being heard in all areas of food safety policy".

Sir Robert May, the government's chief scientific adviser, says that a review of government departments generally shows a marked improvement in the organization and communication of scientific advice within MAFF, although its Spongiform Encephalopathy Advisory Committee has again been criticized for failing to explain sufficiently the scientific evidence underlying its recommendations (see *Nature* 395, 6; 1998). Its chairman, John Pattison, says the committee is "underfunded" and lacks the logistical support needed for proper public communication.

Berlin biomedical centre will boost collaboration

[MUNICH] A new Biomedical Innovation Centre opened last week in Berlin. Some DM40 million (US\$24 million) for its construction were provided by the European Union's Fund for Regional Development and by Germany's federal and state governments.

The centre is to provide low-rent laboratory and office space for biotechnological start-up companies focusing on medical applications. The companies will cooperate closely with scientists at the nearby Max-Delbrück Centre for Molecular Medicine, one of Germany's 16 national research centres. The new buildings have already been let to 21 young companies, with almost 300 employees expected by the end of the year.

UK project puts focus on the science of ageing

[LONDON] The British government last week launched a series of research projects designed to improve the quality of life of Britain's 10 million people over the age of 65. The £5 million (US\$8.4 million) Science of Ageing project consists of 29 research activities, ranging from the early diagnosis of age-related illnesses, such as dementia, to the development of treatments to slow down the process of ageing.

Mark Ferguson, professor of biological sciences at the University of Manchester, said that one of the project's broad aims was to identify as early as possible those individuals prone to an age-related illness, and thus reduce their chances of the disease.

Britain tightens rules to protect natural habitats

[LONDON] The British government has proposed tough new measures to protect natural habitats by denying landowners the right to demand compensation for not using the sites known as Sites of Special Scientific Interest (SSSIs). In future, landowners will only be paid by the government if they agree to manage the sites for the benefit of conservation.

There are nearly 5,000 SSSIs, two-thirds of which are privately owned. The sites are identified because of their distinguishing flora, fauna or geographical features. "Instead of paying landowners for profit foregone, with no guarantee that the value of the sites will be preserved, we will be using public money to ensure proper management," said Michael Meacher, the environment minister.

India launches drive to become IT superpower

[NEW DELHI] The Indian government has approved a US\$10 million budget to set up the first in a series of national institutes of information technology (IT). These are intended to generate the massive manpower it says is needed to achieve the goal of making the country an 'information technology superpower' within ten years.

Science minister Murli Manohar Joshi announced last week that the first IT institute, to be set up in Allahabad in Uttar Pradesh state, will enrol students for a five-year course from July 1999. As well as being a centre of excellence in research and development in IT and allied sciences, the institute "will develop functional links with industries of national and international repute for sustained growth and development without depending on government funding".

US action to curb noise threat to sea mammals

[BOSTON] The US National Marine Fisheries Service is to develop guidelines to protect marine mammals from acoustic pollution from human sources. Experts met at the agency's headquarters in Silver Spring, Maryland, last week to help provide a scientific basis for the proposed guidelines.

"Acoustic pollution is a newly recognized

form of harassment that can have serious effects on these animals," says Roger Gentry, an animal behaviourist in the Office of Protected Resources at the agency. "We can't regulate the volume of noise, *per se*, but if the sound you produce is likely to disturb marine mammals, you would first need authorization."

Draft guidelines should be ready by next spring, and final rules a year or so later. "Given that the science is not too far along, any guidelines would have to be interim, updated with the latest scientific information," says Peter Tyack, a biologist at the Woods Hole Oceanographic Institution, Massachusetts.

China seeks ways to limit future flooding

[LONDON] China's scientists received a mild rebuke from Li Peng, chairman of the National People's Congress, for the failure of the country's one-year-old flood prevention law to control last month's floods.

At a meeting of the National People's Congress this month scientists said that the law was essentially sound but had not been properly enforced. They urged stricter compliance, including punishment for those found breaking the law. They also called for greater checks on soil erosion, and more planting of trees and grass.

But Li said in his closing remarks that flood prevention plans for major rivers needed revision, and that measures needed to be taken as soon as possible. These included setting up a system to support the flood water storing zone, as well as the launching of flood insurance.

Meteoroid dust blanket shrouds Martian moon



[WASHINGTON] Recent data from the Mars Global Surveyor spacecraft suggest that the Martian moon Phobos is covered by a blanket of powdery dust at least a metre thick. The surveyor's thermal emission

spectrometer measured extreme temperature variations between the sunlit and dark sides of Phobos, indicating rapid heat loss by very fine dust particles as the moon rotates every seven hours.

Close-up images of the large crater Stickney (above, upper left) also show light and dark streaks on the crater walls, probably the mark of debris slides on the dusty slopes. The powdery soil of Phobos is the result of millions of years of bombardment by meteoroids.

First shots fired in biological warfare

Sir—Redmond *et al.* have described in Scientific Correspondence¹ new evidence for a covert German attempt to attack Norwegian reindeer with biological weapons to interfere with their use as draught animals to ferry British supplies across northern Norway to Russia during the First World War. These events were not an isolated incident but were part of an ambitious programme of biological warfare directed against animals in neutral trading partners of the Allied forces from 1915 to 1918.

Secret agents were sent to at least five countries (Romania, Spain, Norway, the United States and Argentina) with microbial cultures and instructions to infect shipments to the Allies of horses, mules, cattle and sheep. The bacteria used were those that cause anthrax and glanders^{2–4}.

Germany's programme of biological sabotage began in 1915, with a series of attacks on animals on the eastern seaboard of the United States (neutral until 1917), whence were shipped large amounts of material, including horses and mules, to the Allies. The central player in the US biological sabotage programme, and possibly its architect, was Anton Dilger, an American raised in Germany and trained as a physician there. He returned to the United States in early 1915 and set up a basement laboratory in Washington DC, where he grew cultures of anthrax and glanders. The microbes were

suspended in liquid in test-tubes, and a crew of longshoremen recruited by the Germans wandered among the stockades where animals were collected for trans-shipment, jabbing them with needles dipped into the microbial cultures. This went on for about a year, until a few months after Dilger returned to Germany early in 1916.

In mid-1915, Captain Rudolf Nadolny of the general staff's Berlin headquarters (probably Dilger's boss) shipped anthrax and glanders cultures to the German embassy in Bucharest for Bulgarian agents collaborating with the Germans, targeting the Romanian animal trade with Russia. The programme came to a halt in August 1916, when Romania broke its neutrality and declared war on Austria-Hungary. After the Central Powers' diplomats were expelled, the Romanian police searched the grounds of the German legation, discovering anthrax and glanders cultures. No-one at the time suspected that it was evidence not merely of intent to commit biological warfare, but rather of a continuing operation.

Biological sabotage began in neutral Spain a bit later, although the details are much less well documented. It appears that Spanish horses to be shipped to France were the main targets, although other targets in the French Pyrenees and in Portugal were probably also involved. The Spanish

programme, and its offspring in Argentina, shared with the Norwegian programme the use of ampoules of bacteria concealed in sugar cubes to be fed to intended victims.

Spain was the staging point for shipment of cultures and agents to neutral Argentina, a major supplier of cattle, horses and mules to the Allies. A German secret agent, Herman Wuppermann, travelled by U-boat (probably carrying cultures) from Croatia to neutral Spain, then by commercial steamship to Argentina. He apparently did not establish his own culture lab in Buenos Aires, and was dependent on replenishment with microbes in sugar cubes from Berlin.

The Argentinian programme appears to have ended in 1918, a victim of the increasing difficulty of international transport of microbes and agents. So ended the first modern use of microbes as weapons, and the only documented instance of deliberate attack on neutral countries with microbial agents. Its effect has yet to be adequately evaluated.

Mark Wheelis

Section of Microbiology, University of California, 1 Shields Avenue, Davis, California 95616, USA

1. Redmond, C. *et al.* *Nature* **393**, 747–748 (1998).
2. Wheelis, M. in *Biological Warfare From the Middle Ages to 1945* (eds Geissler, E. & Moon, J. E. v. c.) (Oxford Univ. Press, New York, in the press).
3. Geissler, E. *Militärgeschichtliche Mitteilungen* **56**, 107–155 (1997).
4. Geissler, E. *Biologische Waffen — nicht in Hitlers Arsenalen: Biologische und Toxinkampfmittel in Deutschland 1915–1945* (LIT, Münster, 1998).

How Diana climbed the ratings at the zoo

Sir—“That which we call a rose by any other name would smell as sweet”, but it seems that our perceptions of animals are affected by their names in interesting ways.

Visitors to London Zoo were asked to rank eight photographs of animals in the order that they would choose to help their conservation. The study included 57 taxa and was designed so that, after 57 respondents, each picture had been displayed with every other exactly once, and had appeared once in each position on the display. The design was repeated, with different visitors, but with common names added as captions to the pictures.

The two sets of rankings showed good agreement in general ($r > 0.8$), with the ‘named’ and ‘unnamed’ ranks of most species being within six places. In common with other surveys, we found that big cats were extremely popular choices: the Sumatran tiger and Persian leopard topped both charts, with the Asiatic lion always in

Table 1 **Species showing most difference in rankings between the two surveys**

Highest climbers			Biggest fallers		
Species	Rank without names	Rank with names	Species	Rank without names	Rank with names
Diana monkey	18 =	6 =	Fennec fox	7	21 =
Rothschild's mynah/Bali starling	28 =	8 =	Red-faced black spider monkey	10 =	26 =
Royal python	48	24 =	Strawberry poison frog	23 =	39 =
British warbler cricket	53 =	26 =	Southern tomato frog	32	45
Red-bellied piranha	49	30 =	Hyacinth macaw	21 =	46
Pallid gerbil	43	30 =	Rodrigues fruit bat	36 =	50
			Green imperial pigeon	34	51 =
			Red-back black widow spider	35	56

the top five. Polynesian tree snails were a fixture in last place.

In contrast, the name had a marked effect on the positions of several species (Table 1). Names including words with obvious negative connotations (such as poison, spider, bat) seemed to affect species' rankings adversely. Inspection of species whose rankings moved the other way suggests unexpected insights into the national psyche — the highest climber was also the only species whose name indicated it to be British. And the Diana monkey rose by 12 places into the top ten when named. Interestingly, this

climb is only half that managed by the royal python — and indeed only on a par with the performance of the pallid gerbil.

Claire Carvell

143 Bouverie Avenue South, Salisbury SP2 8EB, UK

Nicholas F. J. Inglis

Dept of Mathematics, Sultan Qaboos University, PO Box 36, Al-Khod, Post Code 123, Sultanate of Oman

Georgina M. Mace

Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

Andy Purvis

Department of Biology, Imperial College, Silwood Park, Ascot SL5 7PY, UK

Worm holes and avian space – time

Kathryn Jeffery and John O'Keefe

Can animals remember where and when events happened? A study of birds that hoard and then retrieve their food shows that they can, and may ultimately provide clues about how human memories are formed.

How do animal memories resemble human memories? This is an important question, because commonalities of memory in species whose evolution diverged aeons ago may point to the fundamental mechanisms of information storage in the brain. Birds, for example, are a class of vertebrates whose last shared ancestor with humans lived about 250 million years ago. Now, on page 272 of this issue, Clayton and Dickinson¹ show that one species of bird can, like humans, remember not only what happened, and where, but also *when* it happened. Unravelling the neural substrate of this memory for time in birds will therefore not only illuminate studies of animal behaviour, but also help to solve a problem in the field of human memory — where and how is memory for events (episodic memory) formed and stored?

The 'what' and 'where' of memory in birds has already been explored in an elegant series of ecologically motivated experiments. These studies capitalized on the ability of some species of bird to hoard food in many locations, in separate caches, to see them through the winter. Food-storing birds such as parids (titmice and chickadees) and corvids (jays, nutcrackers and magpies) readily do this in the laboratory, so the birds can be watched under controlled conditions to find out what information they use to retrieve their food caches.

A typical experiment involves allowing the birds to hoard food — either freely, or within experimentally designated sites in an enclosure — and then manipulating the caches before the birds go back to find them, to work out how the birds know where the food is hidden. Such research has revealed that the birds do not use sensory characteristics of the caches, such as their appearance or smell, but remember instead their spatial locations². This requires a phenomenal memory capacity, and some birds can remember the whereabouts of more than a thousand caches. The neural substrate of this spatial memory is a region of the dorso-medial bird brain that shares embryological and anatomical features with the mammalian hippocampus — a structure that has long been implicated in spatial memory in mammals³.

As well as remembering where they have stored food, birds can remember what they have stored. For example, black-capped chickadees recover their favourite food first and their least favourite food last⁴. Scrub jays that are fed to satiety on one type of food will retrieve an alternative type first⁵. Clayton and Dickinson¹ have now used the 'what' and 'where' abilities of these birds in an ingenious attempt to find out whether they can also remember when they stored the food.

Using the fact that birds retrieve their favourite food first, Clayton and Dickinson

introduced time into the equation by allowing scrub jays (*Aphelocoma coerulescens*) to cache wax worms (Fig. 1). These are a favourite of the birds when fresh, but, in time, they perish and become unpalatable. So, wax worms are the birds' preferred food if retrieved soon after caching, but least preferred if they are retrieved some time after. If the birds can remember when they cached the worms, then they should retrieve them first if they were stored recently and last if they were stored some time ago.

This is exactly what happened. Birds given a choice between retrieving worms or retrieving peanuts chose the worms only if they had been recently cached (four hours earlier), and the peanuts if the worms had been cached a long time ago (five days earlier). Only birds that had previously learned that the worms degraded over time avoided the old caches, showing that this behaviour is not hard-wired. Moreover, Clayton and Dickinson also ruled out the possibility that the birds were using smell to identify the rotten worms.

In a second experiment, the value of old worm caches was degraded in a different way — by being pilfered (removed) by the researchers after five days, but not after four hours. Again, the birds showed a strong tendency to prefer the less tasty but more reliable peanut caches if the worms had been stored a long time ago. Thus, the birds could clearly remember not only what they had stored in which cache, but also whether or not it was stored recently.

How is the 'what', 'where' and 'when' of the food caches stored in the birds' brains? One possibility is that, at the time of caching, the birds form an internal map of the cache locations. Each location is 'labelled' with the contents and approximate time of storage, as well as other information (such as likelihood of the cache being pilfered). This would allow a bird to use the most efficient retrieval strategy to maximize its chances of survival during the winter (for example, by visiting the perishable caches early on, or by varying the types of food retrieved to obtain a range of nutrients). The labels attached to each cache location might be allocated to different regions of the brain, depending on the type of information — the contents of the cache in one region, the time of storage in a separate area that is specialized for temporal representation, and so on.

The experimental model used by Clayton and Dickinson provides a powerful tool for investigating these phenomena further. For example, can the birds remember the order in which different caches were formed, or merely whether each was made recently or some time ago? How fine-grained is the representation (for example, can the birds discriminate five days ago from three days ago)? What brain regions are used to store these different types of information? ►



Figure 1 **Memorable meal** — scrub jay (*Aphelocoma coerulescens*) hiding food in the laboratory.



100 YEARS AGO

Our present knowledge of the theory of errors receives an interesting addition at the hands of M. Charles Lagrange in the form of a contribution to the *Bulletin de l'Académie royale de Belgique* (vol. xxxv. part 6). Without going into details of a purely mathematical nature, certain of M. Lagrange's conclusions appear to be sufficiently important to be worth noticing. In taking the arithmetic mean of a number of observations as the most probable value of the observed quantity, common sense suggests that any observations differing very widely from the rest should be left out of count as being purely accidental, and thus likely to vitiate the result. But as it is impossible to draw the line from theoretical considerations between values retained and values omitted, any such omission would necessarily be unjustifiable. This discrepancy between theory and common sense is, to a large extent, reconciled by M. Lagrange's "theory of recurring means." According to this theory, the *weight* to be assigned to any observation is inversely proportional to the square of the error of the observed value relative to the most probable value. ... The weighted mean is then taken as a second approximation to the most probable value. This mean determines a fresh series of weights to be assigned to the observations by which a new weighted mean ... is found, and so on ... These successive means are called by M. Lagrange "recurring means," and by their use the effects of sporadic errors are, to all practical purposes, eliminated, since the weight assigned to the corresponding observations soon becomes relatively small.

From *Nature* 15 September 1898.

50 YEARS AGO

In the possession of the Science Museum, London, there are six beautifully engraved buttons, classified as diffraction gratings, which are still regarded as masterpieces. They were the work of Sir John Barton, deputy comptroller of the Royal Mint in the early part of the nineteenth century, about whom little is known personally, but who must have been an ingenious inventor and capable engineer, for in 1806 he invented a differential screw measuring instrument capable of measuring 10^{-5} inch. From *Nature* 18 September 1948.

These findings are significant because of the growing interest in the mechanisms that underlie the formation of episodic memory. Episodic memory is memory for events⁶, each of which occurs in a unique setting of space and time. As such, it can be distinguished from semantic memory, which is memory for facts, or 'knowledge'. Episodic memory is disproportionately affected in some types of amnesia, such as that seen in Alzheimer's dementia, and it is thought to depend on the hippocampus⁷, an important structure for spatial memory in both birds and mammals. To store an episodic memory, some method is needed for temporally ordering the sequence of happenings that make up an event. Perhaps episodes can be temporally ordered in the episodic memory of non-humans, as well as humans. In this light, the finding that animals as different

from us as birds possess a mechanism for representing spatiotemporal events is enormously important, and could be a big step towards understanding how space, time and events are represented and remembered in the vertebrate brain. □

Kathryn Jeffery and John O'Keefe are in the Department of Anatomy and Developmental Biology, University College London, Gower Street, London WC1E 6BT, UK.
e-mails: j.okeefe@ucl.ac.uk
kate@maze.ucl.ac.uk

1. Clayton, N. S. & Dickinson, A. *Nature* **395**, 272–278 (1998).
2. Shettleworth, S. J. & Krebs, J. R. *J. Exp. Psychol. Anim. Behav. Process.* **8**, 354–375 (1982).
3. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, 1978).
4. Sherry, D. F. *Anim. Behav.* **32**, 451–464 (1984).
5. Clayton, N. S. *Neuropharmacology* **37**, 441–452 (1998).
6. Tulving, E. *Organization of Memory* (Academic, New York, 1972).
7. Vargha-Khadem, F. et al. *Science* **277**, 376–380 (1997).

Mathematics

Magic squares cornered

Martin Gardner

Dame Kathleen Ollerenshaw, one of England's national treasures, has solved a long-standing, extremely difficult problem involving the construction and enumeration of a certain type of magic square. The solution comes in a book* written with David Brée.

First, some background on magic squares, and their hierarchy of perfection. For many centuries, mathematicians — especially those concerned with combinatorics — have been challenged by magic squares. These are arrangements of n^2 distinct integers in an $n \times n$ array such that each row, column and main diagonal has the same sum. The sum is called the magic constant, and n is called the square's order. Traditional magic squares are made with consecutive integers starting with 0 or 1. If it starts with 0 it can be changed to a square starting with 1 simply by adding 1 to each cell.

No order-2 square is possible. The order-3 square (Fig. 1) barely exists. Why? Because there are just eight different triplets of distinct digits from 1 to 9 that add up to 15, the square's constant. Each triplet appears as one of the square's eight straight lines of three numbers. The pattern is unique — except for rotations and mirror reflections, which are only trivially different.

This little gem of combinatorial number theory was called the *lo shu* in ancient China, meaning 'Lo River writing'. Legend has it that in the 23rd century BC, a mythical King Yu saw the pattern on the back of a sacred turtle in the River Lo. (Modern historians, however, find

no evidence that the pattern was known before the fourth or fifth century BC.) The Chinese identify it with their familiar yin-yang circle. The even digits, here shown shaded, are linked to the dark yin; the Greek cross of odd digits is linked to the light yang. For centuries the *lo shu* has been used as a charm on jewellery and other objects. Today, large passenger ships often feature the *lo shu* on their main deck as a pattern for the game of shuffleboard.

At order 4, the number of magic squares jumps to 880. Among them is a special subset of 48 squares called pandiagonal, which have three amazing properties. This is illustrated by the example in Fig. 2, whose constant is 30.

First, each broken diagonal also adds up to 30. The sequences 0, 3, 15, 12 and 7, 13, 8, 2 are examples. This can be expressed in another way: imagine an endless array of this square placed side-by-side in all directions to make a wallpaper pattern. Then every 4×4 square drawn on the pattern will also be a pandiagonal magic square — in other words, every straight line of four numbers will add up to 30. Second, every 2×2 square on the wallpaper also adds up to 30. Third, along every diagonal, any two cells separated by one cell add up to 15.

In general, a magic square is called pandiagonal if all its broken diagonals add up to the magic constant. Such squares can be constructed of any odd order above three and of any order that is a multiple of four. If a pandiagonal square also has similar properties to the order-4 pandiagonals, it is called 'most perfect': for example, the most-perfect order-8 square in Fig. 3 has a magic constant of 252, and its 2×2 sub-squares add up to 126, and any two numbers that are $n/2 = 4$

*Most-Perfect Pandiagonal Magic Squares: Their Construction and Enumeration. Due for publication on 1 October by The Institute of Mathematics and its Applications, Catherine Richards House, 16 Nelson Street, Southend-on-Sea, Essex SS1 1EF, UK.

2	7	6
9	5	1
4	3	8

Figure 1 *Lo shu*, the only 3×3 magic square.

0	13	6	11
7	10	1	12
9	4	15	2
14	3	8	5

Figure 2 A 4×4 magic square that is **pandiagonal** — the broken diagonals also add up to its magic constant of 30.

cells apart add up to $n^2 - 1 = 63$.

Although all order-4 pandiagonals have been known to be most perfect for three centuries, little has been known about most-perfect squares of higher orders. There was no method of constructing them all, or even of determining the number of squares of a given order.

These questions are finally settled by Kathleen Ollerenshaw and David Brée in their new book. The authors have devised a method for constructing all the most-perfect squares of any order, and a way of calculating their number.

Unlike the ordinary pandiagonals, there are no most-perfect squares with odd order, so the only possible orders are multiples of four. At each leap in order, the number of essentially different most-perfect squares increases rapidly: from 48 squares of order four, to 368,640 of order eight, to 2.22953×10^{10} of order 12. When you reach order 36, the number is 2.76754×10^{44} — around a thousand times the number of pico-pico-seconds since the Big Bang.

This solution of one of the most frustrating problems in magic-square theory is an achievement that would have been remarkable for a mathematician of any age. In Dame Kathleen's case it is even more remarkable,

0	62	2	60	11	53	9	55
15	49	13	51	4	58	6	
16	46	18	44	27	37	25	
31	33	29	35	20			
52							

Figure 3 A most-perfect square of order eight. Its rows, columns, diagonals and broken diagonals all add up to 252, and all 2×2 sub-squares add up to 126. Kathleen Ollerenshaw and David Brée have found a way to construct most-perfect pandiagonal magic squares of any order.

because she was 85 when she and Brée finally proved the conjectures she had earlier made. In her own words, "The manner in which each successive application of the properties of the binomial coefficients that characterize the Pascal triangle led to the solution will always remain one of the most magical mathematical revelations that I have been fortunate enough

to experience. That this should have been afforded to someone who had, with a few exceptions, been out of active mathematics research for over 40 years will, I hope, encourage others. The delight of discovery is not a privilege reserved solely for the young." □

Martin Gardner is at 3001 Chestnut Road, Hendersonville, North Carolina 28792, USA.

Bacterial infection

For whom the bell tolls

Craig Gerard

"Now this bell tolling softly for another, says to me, Thou must die."

The 1623 Meditation 17 by the English metaphysical poet John Donne was probably inspired by the church bells that tolled to announce death by the plague. Death in this and many other infectious diseases typically follows septic shock, often caused by so-called Gram-negative bacteria. The path that leads from these organisms to septic shock has been under investigation for over a century, and, on page 284 of this issue, Yang *et al.*¹ report that a cell-surface receptor, Toll-like receptor 2 (TLR2), may be part of the long-awaited solution to the puzzle.

In 1884, the Danish physician Christian

Gram discovered that the outer membranes of bacteria may be classified as Gram negative or Gram positive, based on the ability of acetone/alcohol to decolorize cells stained with Gram's iodine and crystal violet. Gram-negative organisms contain a complex glycolipid in their outer membranes, known as lipopolysaccharide (LPS) or endotoxin. In picogram quantities, this substance can induce mammalian white blood cells to secrete cytokines which, if left unchecked, can lead to fever, coagulation defects, lung dysfunction, kidney failure and circulatory collapse.

How are trace quantities of this glycolipid recognized, and the signals necessary for the production of toxic cytokines transduced?

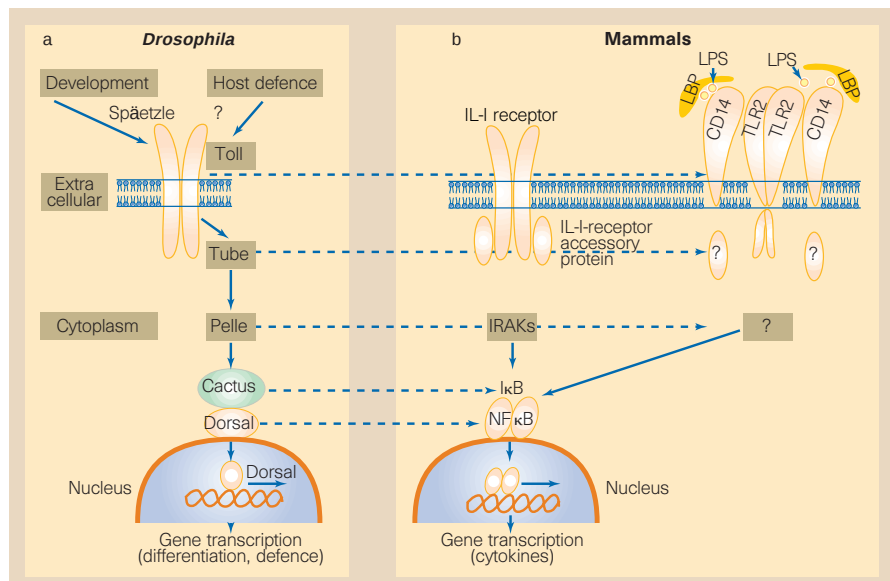


Figure 1 Functions of Toll — from development to innate immunity. a, In *Drosophila*, Toll acts in development and, later, in defence. Antibiotic-gene products, regulated by Dorsal, protect the adult fly. b, Homologues of this system in mammals. Dashed arrows represent directly related mammalian genes in an evolutionary sense. For example, Toll and the interleukin-1 receptor are transmembrane receptors with some structural homology in their signal-transducing intracellular domain. Cactus and Dorsal are direct homologues of mammalian I- κ B and NF- κ B. Yang *et al.*¹ now show that the Toll-like receptor 2 (TLR2) is a member of this ancient innate immune-response system. LPS, lipopolysaccharide; LBP, LPS-binding protein; IRAK, IL-1 receptor accessory protein kinase.

The innate immune response, which is activated in response to potential pathogens, is a primitive form of host defence. Largely non-specific, it consists of 'pattern-recognition' proteins². For example, many pathogens contain repeating structures of carbohydrates and peptide derivatives, as well as methylated DNA, which the host can recognize as 'non-self'. After a molecular structure associated with a potential pathogen is recognized, the innate immune response seeks both to dispose of it and to activate the body's white blood cells, gearing up for a battle. The disposal system is usually engulfment of the pathogen by macrophages, whereas the battle alert is sounded through the production of cytokines.

The pro-inflammatory cytokines — tumour-necrosis factor and interleukin-1 (IL-1), to name but two — nonspecifically activate the host at several levels, to destroy the invading pathogen. But there is some risk. Although too little response favours infection, overreaction initiates toxicity. The lethal effects of the cytokines underscore the dichotomy between host defence and the inflammatory response.

So how does the innate immune system recognize endotoxin shed from bacteria? The LPS-binding protein (LBP) is a circulating, high-affinity binding site for LPS. Endotoxin naturally aggregates, and LBP breaks down these aggregates for presentation to a second molecule, CD14 (refs 3, 4). Here the mystery begins. CD14 is tethered to the host-cell membrane by a glycosyl-phosphatidylinositol anchor — it is not intrinsically con-

nected to the transmembrane signal-transduction apparatus⁵. Thus, although LBP and CD14 are primary in recognizing LPS, they must present endotoxin to an elusive co-receptor that initiates signal transduction. Much has already been discovered about this signalling pathway: it requires certain protein-kinase cascades, protein kinase C, activation of the transcription factor NF- κ B and transcription of cytokine genes⁵.

One year ago, Charles Janeway and colleagues found⁶ a human homologue of a *Drosophila* receptor protein called Toll on cells of the immune system. This protein could activate signal transduction, leading to the production of pro-inflammatory cytokines. Toll was initially identified as a key protein in development of the *Drosophila* body plan. Later in a fly's life, this same protein is involved in the defensive response to fungal infection. Toll is a transmembrane protein, and the intracellular portion contains a motif associated with the receptor for IL-1 (ref. 7). The Janeway team posited that the human Toll-receptor homologue might be an important adaptor receptor, bridging innate immune recognition with signal transduction. Five human Toll-like receptors (TLR1–5) have subsequently been cloned⁸.

Yang and colleagues¹ now show that human epithelial cells engineered to express TLR2 can respond, sensitively and selectively, to LPS in the presence of LBP and CD14. The response involves activation of genes in the cytokine cascade, particularly NF- κ B. Wild-type (but not chemically modified) LPS binds specifically, albeit with a low

affinity, to a TLR2 extracellular leucine-rich repeat segment. Thus, the purpose of LBP and CD14 may be to present disaggregated LPS to a Toll receptor. This, in turn, could promote dimerization of the Toll receptor and recruitment of the intracellular signal-transduction complex.

Yang *et al.* also show that the intracellular region at the extreme carboxy terminus of TLR2 is essential for signal transduction. Moreover, this region has some homology to the receptor for IL-1. Signal transduction by *Drosophila* Toll and the IL-1 receptor involves dimerization of the receptor, driven by the ligand, then recruitment of a number of associated intracellular serine/threonine and tyrosine protein kinases. These kinases activate the signal-transduction cascade, including mitogen-activated protein kinases and NF- κ B. Thus, the model in Fig. 1 seems to be a common innate-immune mechanism that extends from flies to man.

Curiously, a strain of mice that is resistant to the effects of LPS has also been described⁹. This resistance has been mapped to a single, autosomal locus on mouse chromosome 4. The genetically conserved equivalent region in the human genome is chromosome 9q32–33, which turns out to be the locus for the first reported Toll homologue, TLR4. It is tempting to speculate whether the LPS locus similarly encodes a Toll homologue, or a related adaptor protein.

In any event, the identification of a candidate gene for LPS signal transduction could lead to new approaches for the treatment of endotoxin/LPS-associated shock. Knowing the appropriate extracellular receptor and intracellular adaptors could facilitate attempts to develop antagonists to this lethal syndrome. But a word of caution. Yang and colleagues' results are derived from transfected cells alone — not native human cells bearing TLR2. Furthermore, the modest effect of CD14 is not entirely consistent with the much bigger part it plays in models *in vitro* and *in vivo*. Nevertheless, the current work should stimulate definitive experiments on the role of TLRs in the response to LPS in the wild. □

Craig Gerard is at the Children's Hospital, Perlmutter Laboratory, Harvard Medical School, Boston, Massachusetts 02115, USA.
e-mail: gerard_c@a1.tch.harvard.edu

1. Yang, R.-B. *et al.* *Nature* **395**, 284–288 (1998).
2. Medzhitov, R. & Janeway, C. A. Jr *Cell* **91**, 295–298 (1997).
3. Schumann, R. R. *et al.* *Science* **249**, 1429–1431 (1990).
4. Wright, S. D., Ramos, R. A., Tobias, P. S., Ulevitch, R. J. & Mathison, J. C. *Science* **249**, 1431–1433 (1990).
5. Ulevitch, R. J. & Tobias, P. S. *Annu. Rev. Immunol.* **13**, 437–457 (1995).
6. Medzhitov, R., Preston-Hurlbut, P. & Janeway, C. A. Jr *Nature* **388**, 394–397 (1997).
7. Belvin, M. P. & Anderson, K. V. *Annu. Rev. Cell Dev. Biol.* **12**, 393–416 (1996).
8. Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A. & Bazan, J. F. *Proc. Natl Acad. Sci. USA* **95**, 588–593 (1998).
9. Watson, J., Riblet, R. & Taylor, B. A. *J. Immunol.* **120**, 422–424 (1977).

Nuclear and subnuclear boiling

Frank Wilczek

Can a nuclear material, compressed hard enough, turn into a still denser substance known as quark matter? According to Stephanov, Rajagopal and Shuryak of Stony Brook and MIT, we may be able to find out by doing experiments with existing particle accelerators¹. They apply the theory of phase changes to extend a bold analogy originally made by Niels Bohr.

Bohr's liquid-drop model of the nucleus is a crude but useful tool for understanding the masses and behaviour of heavy nuclei. The idea is that both liquids and atomic nuclei consist of closely but not rigidly packed hard constituents — molecules in one case, protons and neutrons in the other — so their equilibria and collective motions should follow similar patterns. For example, charged water drops are unstable to fission: they tend to break into smaller droplets. This was discussed theoretically by Lord Rayleigh in the nineteenth century and investigated experimentally by Bohr in his thesis. Many years later, when nuclear fission was discovered, Bohr quickly realized the analogy, and was able to carry over much of the old analysis². At the moment, a simple but startling question suggested by the liquid-drop analogy is receiving a lot of theoretical and experimental attention³. Does nuclear matter, like water, boil when heated?

On first hearing, it might seem that this question takes a model that was meant to be tongue-in-cheek far too literally. Upon reflection, however, one begins to realize its depth. For the logic of the liquid-drop analogy, rather than being strained, becomes more compelling when nuclei are heated. Indeed, a

great limitation of the liquid-drop picture is that for nuclei near their ground state, but not for ordinary liquids, quantum effects are important. Roughly speaking, the possible motions of the protons and neutrons are constrained, because they are allowed only certain discrete energies. There is no such constraint for ordinary liquids.

But the constraints introduced by quantum mechanics become less important — and the logic of the liquid-drop picture more convincing — when the nuclei are heated, so that there are many quanta of energy to share. Moreover, many important features of transitions are quite independent of the microscopic details, or 'universal'⁴. In particular, when the relevant control parameters, such as temperature and density, are close to those for which a continuous (second-order) transition takes place, fluctuations between the phases occur even over large distance scales. The nature of these fluctuations does not depend on the details of the underlying dynamics at short distances. It is the same, after renaming the variables, for wildly different systems.

Having convinced ourselves that the question is respectable, we can better appreciate the emerging answer. It is 'yes'. Nuclei do indeed have a sharp boiling phase transition⁵. The results of many measurements (Fig. 1) show how the temperature of nuclei rises as a function of the energy deposited. One sees that, as for water at the boiling point, there is a substantial range where the temperature changes very little in response to applied heat. This is the phenomenon of latent heat, the classic signal of a discontinu-

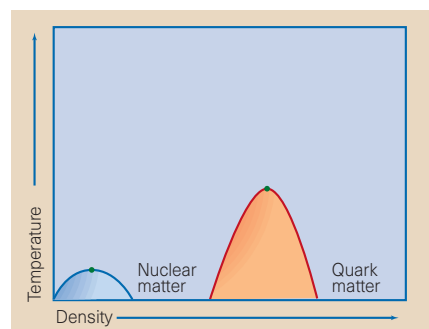


Figure 2 A conjectured phase diagram of extremely dense matter. In the blue region, free nucleons and nuclear matter can coexist in equilibrium, and as nuclear matter boils it passes from right to left through this region. Similarly, if quark matter can exist, there will be a region where nuclear matter and quark matter can coexist in equilibrium (red). The critical points (green) mark the limits of these phase transitions.

ous (first-order) transition, here occurring between the usual nuclear liquid phase and a gaseous phase of dissociated protons and neutrons.

More insight into the physics can be gained by nuclear collisions. Crucially, collisions can produce fireballs with different initial densities, as well as different temperatures. At high density, as we have seen, there is a first-order transition as the temperature is varied, with absorption of latent heat when one passes to the less ordered phase. As the density is lowered, the liquid state is more tenuous, and we would expect the quantity of latent heat to decrease. Then, at some critical density, the transition (as a function of temperature) will occur with zero latent heat — it will be a second-order transition. At still lower densities, the liquid and gaseous states will be continuous.

All this is familiar in the context of liquid–gas transitions, but what does it mean for nuclei? If the proposed phase diagram (Fig. 2) is correct, we should see large fluctuation phenomena in collisions where the nuclear fireball drops out of equilibrium at a temperature and density near the critical point. Specifically, one should see a power-law distribution in the size of fragments. According to universality, the slope in this distribution is precisely related to the distribution of domain sizes in an Ising-model magnet, or to the spectrum of critical opalescence at other liquid–gas transitions. Remarkably, there is some experimental evidence for such a distribution, and therefore for nuclear boiling⁶.

But Stephanov *et al.*¹ are concerned with the possibility of a second phase transition in the opposite direction, when nuclear matter is compressed. Some model calculations^{7,8} suggest that when one squeezes nuclear matter very hard, doing work of around 200–400 MeV per nucleon, there is a first-

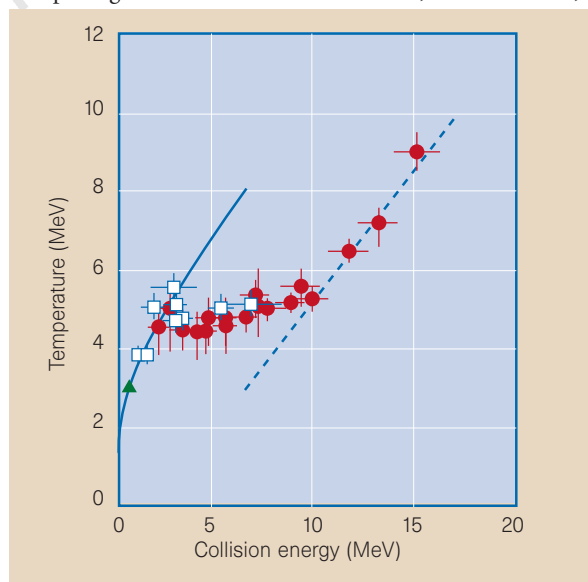


Figure 1 The change in temperature (measured through the kinetic energy of escaping particles) as a function of energy deposited in nuclear collisions. The plateau, where a higher energy input does not raise the temperature, is a sure sign of a discontinuous phase transition. This is nuclear boiling. (Adapted from ref. 5.)

order transition from nuclear matter to a new, sharply distinct phase that is several times as dense — quark matter. The cores of neutron stars may well consist of quark matter.

But this is controversial, and the physical characteristics of quark matter are mostly unknown. Quantum chromodynamics (QCD) is the theory that describes quark interactions, and its equations are very difficult to solve at high densities; even the numerical techniques that are wonderfully successful at low density fail miserably. So experimental information on this regime would be most welcome.

Unfortunately, to achieve sufficient compression to produce quark matter in the laboratory, one needs very energetic collisions — which will also deposit a lot of heat. That makes it very difficult to reach the putative transition at zero temperature. More generally, heat will be released as nuclear matter is compressed, warming the matter and making it expand, driving the material along the coexistence curve in the temperature/density plane towards the critical point.

For that reason, and others, the critical point appears to be a more hopeful experimental target. Stephanov *et al.* propose several ways to identify the subnuclear boiling critical point. Long-wavelength fluctuations should eventually materialize as low-momentum pions, which should be anomalously abundant (and exhibit anomalously

large event-by-event fluctuations in number) near the critical density and temperature. Also, because of the large specific heat near the critical point, collisions with a range of energies should be clustered around the critical temperature.

Finding the position of the critical point would shed light on aspects of QCD that are both fundamental and obscure. (Is quark matter truly a distinct phase? Does it boil into nuclear matter?) It would also be a truly remarkable example of the power of theoretical reasoning and the concept of universality in phase transitions. Hitherto, discussions of high-energy nuclear collisions have been dominated by the quest to create a quark-gluon plasma⁹. Now we see that there is another, equally worthy, goal: to smoke out the transition to quark matter. □

Frank Wilczek is in the School of Natural Sciences, Institute for Advanced Study, Olden Lane, Princeton, New Jersey 08540, USA.
e-mail: wilczek@ias.edu

1. Stephanov, M., Rajagopal, K. & Shuryak, E. preprint hep-ph/9806219 on xxx.lanl.gov
2. Bohr, N. & Wheeler, J. *Phys. Rev.* **56**, 426–450 (1939).
3. Trautmann, W. preprint nucl-ex/9611002 on xxx.lanl.gov
4. Pfeuty, P. & Toulouse, G. *Introduction to the Renormalization Group and to Critical Phenomena* (Wiley, London, 1976).
5. Pochodzalla, J. *et al.* *Phys. Rev. Lett.* **75**, 1040–1043 (1995).
6. Gilkes, M. L. *et al.* *Phys. Rev. Lett.* **73**, 1590–1593 (1994).
7. Alford, M., Rajagopal, K. & Wilczek, F. *Phys. Lett. B* **422**, 247–256 (1998).
8. Rapp, R., Schafer, T., Shuryak, E. & Velkovsky, M. *Phys. Rev. Lett.* **81**, 53–56 (1998).
9. Wilczek, F. *Nature* **391**, 330–331 (1998).

Mutagenesis

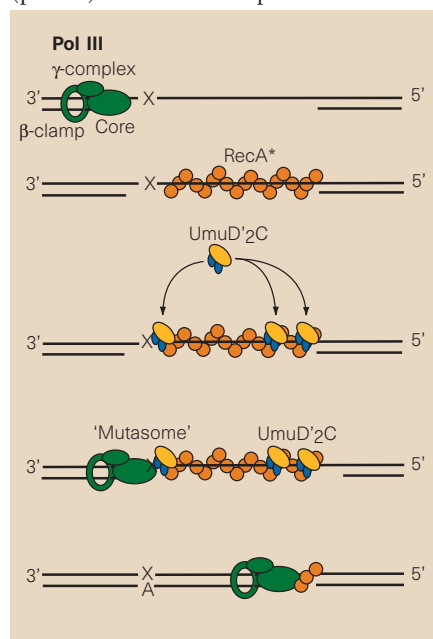
Purposeful mutations

Myron F. Goodman

When DNA is passed from mother to daughter cells during replication, several mechanisms ensure that this occurs with a minimum of mutations. These include proofreading exonucleases and post-replication mismatch-repair enzymes¹. Nevertheless, “there’s many a slip ‘twixt the cup and the lip”², such as the spontaneous loss of DNA bases and DNA damage from external sources. Many cells contain systems to repair damaged DNA, but *Escherichia coli* also harbours proteins that do the opposite — they reduce replication fidelity. These proteins are induced in response to DNA damage as part of the so-called SOS response³, and reports by Bacher Reuven *et al.*⁴ in *Molecular Cell* and Tang *et al.*⁵ in *Proceedings of the National Academy of Sciences* detail the biochemical reconstitution of this system.

Given a choice between mutation and death, perhaps the SOS mutagenic response is a cell’s last ditch attempt to survive, when DNA damage outstrips its repair capacity. Two proteins, UmuD’ and UmuC, are induced in the SOS response to DNA damage by ultraviolet light and most chemical muta-

gens. A trimer of one UmuC and two UmuD’ molecules (UmuD’₂C) works with the replicative DNA polymerase III holoenzyme (pol III) and the RecA protein to form a



‘mutasomal’ complex¹ (Fig. 1). Normally, pol III stalls when it encounters DNA damage, but the mutasomal complex allows pol III to copy and bypass a variety of DNA lesions, including abasic sites, cyclobutane dimers and 6–4 photoproducts.

Since the mid-1970s, when the phenomenon of the SOS response was formalized⁶, researchers have tried to understand the biochemistry behind it. Harrison Echols and co-workers were the first to report⁷ the UmuD’₂C-dependent lesion bypass *in vitro*, using a denatured/renatured form of purified UmuC. But a stumbling block towards reconstituting a reproducible lesion-bypass assay⁸ was the inability to purify UmuC protein in an aqueous soluble form. Now, taking completely different approaches, Bacher Reuven *et al.*⁴ and Tang *et al.*⁵ have overcome this difficulty. Whereas Bacher Reuven *et al.* purified soluble UmuC by fusing it to a portion of the maltose-binding protein, Tang *et al.* purified an intact, native UmuD’₂C complex. Both systems illustrate the remarkable effectiveness of the soluble forms of UmuD’ and UmuC in facilitating lesion bypass.

Bacher Reuven *et al.* used a circular duplex DNA template containing a synthetic tetrahydrofuran abasic lesion located within a 350-nucleotide single-stranded gap. They worked out the sequence of the lesion that had been replicated *in vitro* from the replication products. These were amplified by the polymerase chain reaction, then cloned into plasmids and transfected into *E. coli*. The authors found that in the absence of UmuC, UmuD’, RecA and the single-stranded DNA binding protein (SSB), pol III pauses and then skips over the lesion, generating a one-base deletion. But in the presence of these proteins, lesion bypass is stimulated. In this case, however, the mutagenic specificity of pol III is altered. This results in base substitu-

Figure 1 Possible mechanism of SOS-dependent translesion replication. The replicative DNA polymerase III holoenzyme (pol III), containing a catalytic core, β sliding clamp and γ -complex clamp loader, stalls when it encounters an abasic template lesion (X). RecA protein binds to the region of single-stranded DNA downstream from the lesion, forming an activated RecA* filament. The UmuD’₂C complex binds to various sites along the RecA* filament, but only those bound to the end of the filament are appropriately positioned to form a ‘mutasome’ consisting of UmuD’₂C, RecA*, pol III and, possibly, single-stranded DNA-binding protein (not shown). The nucleoprotein aggregate allows replication to proceed past the lesion, generating a base-substitution mutation, with preference for incorporating dAMP (A) opposite the abasic site. This process can now be dissected biochemically thanks to reconstituted *in vitro* systems developed by Bacher Reuven *et al.*⁴ and Tang *et al.*⁵. (Figure courtesy of R. Woodgate.)

tions and favours incorporation of dAMP opposite the lesion, in agreement with mutational data⁹.

Tang *et al.* carried out replication using a primed bacteriophage M13 DNA template containing a single, synthetic, abasic lesion. They used gel electrophoresis to analyse ³²P-labelled primer-elongation bands. These authors found that synthesis by pol III is almost completely blocked one base before the lesion, and that lesion bypass requires the mutasomal proteins. The data indicate that both of the bypass reactions — misincorporation opposite the lesion and incorporation beyond the lesion — are stimulated by UmuD'2C. Moreover, Tang *et al.* observed no lesion bypass when they used a mutant of RecA that is defective in the SOS response¹⁰. The finding that UmuD'2C also diminishes the fidelity of DNA synthesis at template sites that lack lesions is consistent with observations of SOS-induced untargeted mutagenesis³. Thus, this complex might act as a generalized 'elongation' factor, extending kinetically unfavourable primer ends at undamaged template sites as well as those at DNA lesions.

With a reconstituted bypass assay we can measure misincorporation specificities, bypass efficiencies and sequence context effects, and compare these with the genetic data — SOS mutational spectra and bypass efficiencies have been reported for a variety of template lesions in different sequence contexts¹¹. We can also analyse the mechanism of lesion bypass biochemically. For example, where and when do the Umu proteins enter and depart from the replication

fork? And does the activated RecA (RecA*) at the tip of the filament target UmuD'2C to the site of the lesion¹²?

Advances in understanding nucleotide-excision and post-replication mismatch repair — also initially studied in prokaryotes — have provided insight into human diseases such as xeroderma pigmentosum and a variety of familial cancers. By comparison, the main steps in error-prone translesion DNA synthesis await discovery. Perhaps proteins with properties akin to UmuD' and UmuC will also be found in eukaryotes. For instance, a p53-mediated DNA-damage-induced repair process has been reported in human skin cells¹³, possibly analogous to the bacterial SOS response, and lesion bypass has been shown in progesterone-matured *Xenopus* oocytes¹⁴. Indeed, the reconstituted lesion-bypass assay system could be just what's required to identify 'fidelity-challenged' mutasomal proteins in higher organisms. □

Myron F. Goodman is in the Department of Biology, Program in Molecular Biology, Stauffer Hall of Science 172, University of Southern California, Los Angeles, California 90089-1340, USA.
e-mail: mgoodman@rcf.usc.edu

1. Echols, H. & Goodman, M. F. *Annu. Rev. Biochem.* **60**, 477–511 (1991).
2. Talladas Greek Anthology Book 10, epigram 32 (attributed to Homer).
3. Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* 407–522 (ASM Press, Washington DC, 1995).
4. Bacher Reuven, N., Tomer, G. & Livneh, Z. *Mol. Cell* **2**, 191–199 (1998).
5. Tang, M. *et al. Proc. Natl Acad. Sci. USA* **95**, 9755–9760 (1998).
6. Witkin, E. M. *Bacteriol. Rev.* **40**, 869–907 (1976).
7. Rajagopalan, M. *et al. Proc. Natl Acad. Sci. USA* **89**, 10777–10781 (1992).
8. Petit, M.-A. *et al. J. Biol. Chem.* **269**, 23824–23829 (1994).
9. Lawrence, C. W., Borden, A., Banerjee, S. K. & LeClerc, J. E. *Nucleic Acids Res.* **18**, 2153–2157 (1990).
10. Dutreix, M. *et al. J. Bacteriol.* **171**, 2413–2415 (1989).
11. Horsfall, M. J. & Lawrence, C. W. *J. Mol. Biol.* **235**, 465–471 (1994).
12. Sommer, S., Boudsocq, F., Devoret, R. & Bailone, A. *Mol. Microbiol.* **28**, 281–291 (1998).
13. Eller, M. S., Maeda, T., Magnoni, C., Atwal, D. & Gilchrist, B. A. *Proc. Natl Acad. Sci. USA* **94**, 12627–12632 (1997).
14. Oda, N. *et al. Science* **272**, 1644–1646 (1996).

Ribozymes

How to make a nucleotide

Michael P. Robertson and Andrew D. Ellington

The hypothesis of an early 'RNA world' containing RNA (ribozyme) rather than protein catalysts was originally published in these pages¹ in 1986. Since then the story has developed considerably, and the latest instalment appears on page 260 of this issue. There, Unrau and Bartel² describe the isolation of RNA molecules that can

catalyse the synthesis of the nucleotide 4-thiouridine from 4-thiouracil and activated ribose, thus acting as thiouridine synthetase enzymes. The broad significance of this finding is that the reaction belongs to a class of fundamental biochemical processes, those by which the nucleotide constituents of nucleic acids are created.

Looking backwards from a contemporary vantage point, it seems that the RNA component of modern ribosomes, the cell's protein-synthesizing machinery, may itself be a ribozyme and thus a remnant of an RNA world³. Looking forwards from origins, it is plausible that chemically simple, nucleic-acid or non-nucleic-acid replicators gave rise to the raw material that became the RNA world^{4,5}. But there must have been several evolutionary steps bridging the gap between the earliest replicators and the catalytically sophisticated ribozymes, and we know very little about them. The gap has been filled in part by experimental recapitulations of what may have been possible. Researchers have learned how to carry out Darwinian selection in a test tube, and have culled species of nucleic acids that bind other molecules (aptamers), and ribozymes, from nucleic-acid populations with random sequences. Selected RNAs serve as convenient *doppelgängers* of their long-departed counterparts, and may provide clues to the construction and verification of models for the RNA world.

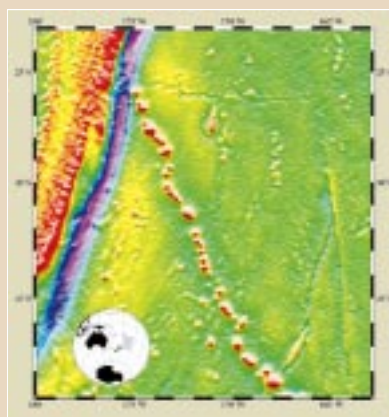
Unrau and Bartel² describe the most sophisticated ribozyme selection to date, the evolution of a ribozyme that can form the glycosidic bond that joins a sugar to a base to

Earth science

Crack-up under the Pacific

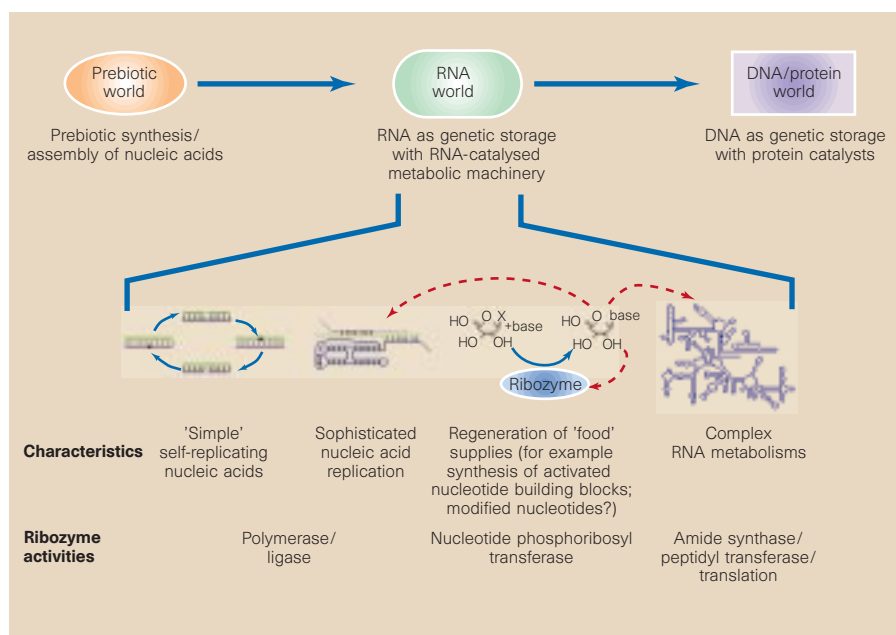
To the left in this gravity map of the ocean floor in the southwest Pacific is the Australian tectonic plate. To the right is the Pacific plate, which is (or should be) plunging under the Australian plate along a wide subduction front that appears in purple. The catch literally comes from the chain of seamounts that appears as bobbles running from the southeast. It is thought that an especially large seamount, some 100 km in diameter and provisionally dubbed Godzilla, has jammed subduction locally.

What, though, of the hair-line crack running due east from the place where the seamount chain meets the subduction front? This is the Louisville trough, and it is in reality about 800 km long and 20 km across. Writing in *Geology* (26, 795–798; 1998), Christopher Small and Dallas Abbott consider two explanations for the trough's origin. One, proposed by Peter Lonsdale, is that it is an extinct spreading



centre, a ridge along which magma once welled up and created new ocean floor. The other, which they put forward, is that the stresses that have built up on the locked Pacific plate may be cracking it — which if correct, as the authors themselves point out, begs the question of why the seamount did not give first.

Tim Lincoln



make a nucleotide. This demonstration fills a critical niche in a generalized scheme for the evolution of the RNA world, as depicted in Fig. 1.

Previously selected ribozymes have been able to manipulate phosphoester and pyrophosphate bonds. They could thus conceivably have built or activated mono- and oligonucleotide substrates necessary for their replication. There would, however, have come a point when any nucleotides in a primordial soup would have been effectively scavenged by early ribo-organisms, and would have had to have been replaced from chemically simpler substances. Unrau and Bartel's selection of thiouridine synthetases

suggests that ribozymes might have existed that could reconstruct nucleoside and nucleotide precursors from compounds with molecular weights as small as 100 daltons. The ribozymes themselves are extremely impressive, with rate enhancements (k_{cat}/K_m) of over ten million times that of the uncatalysed reaction.

The thiouridine synthetases may also represent a chemical innovation that is of more technical interest. As described in the box, protein uridine synthetases act by an S_N1 reaction mechanism, stabilizing a carbocation intermediate. So the RNA thiouridine synthetases are either the first ribozymes to use this type of chemistry or are the first uri-

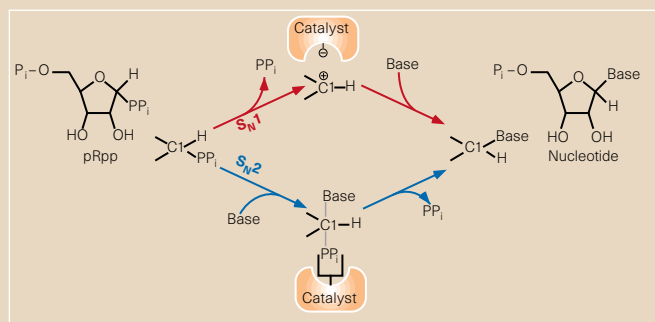
Figure 1 Biochemical epochs in the RNA world. Early nucleic-acid or non-nucleic-acid replicators gave rise to faster and more faithful mononucleotide or polynucleotide polymerases. As foodstuffs for replicators were exhausted, an evolutionary advantage would have accrued to organisms that evolved the ability to generate new building blocks (for instance, using the thiouridine synthetase identified by Unrau and Bartel²). At this stage, ribozymes would have possessed the chemical sophistication to modify nucleotide or oligonucleotide precursors. Modified nucleotides could have improved all extant catalysts and fostered the evolution of more sophisticated catalysts, such as ribosomal RNA. The advent of neither cells nor energy metabolism is explicitly indicated, as either innovation would have yielded an evolutionary advantage irrespective of when it occurred.

dine synthetases to use an S_N2 mechanism.

Just as significant as what the ribozyme thiouridine synthetases tell us about the RNA world is what they don't (and can't). It is possible that their sequences and structures are similar to or even the same as those of analogous enzymes in the RNA world. But unless a ribo-organism is discovered at the Martian polar cap, or in the oceans far beneath Europa's icy crust, it will be difficult to confirm any putative similarities. Paradoxically, as more is learned about ribozymes in selection experiments, less can be reliably inferred about the exact natures of their ancient counterparts.

For example, last year we had the report⁶

Routes to the glycosidic bond



The biosynthesis of nucleotides is accomplished through the creation of a glycosidic bond between a ribose phosphate unit (pRpp) and a purine or pyrimidine base, as shown in the figure here. The bond occurs between C1 of the ribose and N9 of a purine or N1 of a pyrimidine. This nucleophilic substitution

reaction can occur by either of two reaction mechanisms, S_N1 or S_N2 . The S_N1 mechanism proceeds initially by the displacement of a leaving group – pyrophosphate (PP_i) in the case of pRpp – resulting in the formation of a positively charged carbon atom at the ribose C1 position. The carbocation

intermediate is then attacked by the nucleophilic nitrogen of the purine (N9) or pyrimidine (N1) base, resulting in the new glycosidic bond. The protein nucleoside phosphoribosyl transferase enzymes that catalyse the critical reaction of nucleotide biosynthesis within the cells of living organisms all use the S_N1 mechanism to perform this chemistry. One way for protein or RNA enzymes to catalyse this type of reaction is by helping to stabilize the carbocation intermediate (for example, by using negatively charged groups within the catalyst). Alternatively, the S_N2 mechanism is initiated by the approach of a

nucleophile. The nucleophile becomes partially bonded to C1 while the pyrophosphate leaving group becomes simultaneously less strongly bonded. In the transition state of the reaction, both the attacking nucleophile and the leaving group are only partially bonded to C1; the purine or pyrimidine base then becomes fully bonded and pyrophosphate is completely displaced. Ways in which catalysts using this type of mechanism may function include alignment of the nucleophile for the proper angle of attack or stabilization of the leaving group.

M.P.R. & A.D.E.

of the selection of truly novel Diels–Alder synthetase ribozymes that contain modified nucleotides (the significance of these Diels–Alder ribozymes is that they can create carbon–carbon bonds, which is presumably a step towards a sophisticated biochemistry). In these and other experiments, the modified nucleotides have been found to be essential for ribozyme catalytic or aptamer activities⁷. The results confirm that chemical augmentation of the canonical nucleotides can greatly influence catalysis, and they support the general notion that early catalysts may have employed modified nucleotides to improve their performance.

However, such results also greatly complicate the picture. In attempting to peer back past the invention of the ribosome and protein primacy, we must now not only attempt to guess which ribozymes catalysed which primitive reactions, but also which of a number of nucleotide modifications may have been present. Moreover, ribozymes such as thiouridine synthetase could potentially have carried out primordial combinatorial chemistry programmes on their own precursors and thus actively participated in their own chemical evolution (Fig. 1). Because each happenstance chemical or enzymatic synthesis of a modified nucleotide may have drastically altered the sequences, structures and functions of ancient ribozymes, firm knowledge of the denizens of the RNA world becomes ever more diffuse.

In this light, attempts to use selection experiments to explain the origin of the genetic code or of ribosomal RNA are unlikely to succeed, and any postulated sequence similarities between selected and ancient RNA molecules are better explained by random chance than by historical necessity⁸. Stephen Gould has argued that re-running the tape of life would yield a vastly different complement of organisms. Similarly, simulations of that tape probably yield very different molecules from those that once existed.

Nonetheless, selection experiments can at least indicate the boundary conditions for early evolution. For example, whereas only three ribozyme families emerged from eleven rounds of selection for a thiouridine synthetase, several were isolated following ten rounds of selection for RNA ligases⁹. Because the ligases carry out simpler chemical reactions, the difference in numbers of ribozyme families isolated supports a principle that we might intuitively expect to be the case — that there is a relationship between informational and chemical complexity.

Similarly, as we learn more about the efficacies of modified nucleotides in selection experiments, we may be able to identify which modifications would have been most helpful to nascent nucleic-acid catalysts, even if we can never know precisely what

those catalysts were. As with the work by Unrau and Bartel, such enquiries should allow us to better see, and remain faithful to, the concept of an RNA world. But they will also increasingly force us to be agnostic as to whether any one model is true. □

Michael P. Robertson and Andrew D. Ellington are in the Department of Chemistry and Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712, USA.
e-mail: andy.ellington@mail.utexas.edu

Signal transduction

Docking I κ B kinases

Claus Scheidereit

In response to stress or inflammatory and immune reactions, mammalian cells rapidly send a transcription factor called NF- κ B into the nucleus. Once there, this protein activates the transcription of specific genes, eliciting various cellular responses. Without the right extracellular signals, however, NF- κ B is kept under tight control in the cytoplasm, coupled to inhibitory I κ B proteins (I κ B α , β and ϵ), which block its transport into the nucleus. NF- κ B is critical for proper immune function, cell growth and survival, and anomalous activation is associated with inflammatory and neoplastic diseases and viral infection.

Central to the activation of NF- κ B are two I κ B kinases (IKKs), and, on pages 292 and 297 of this issue, Cohen *et al.*¹ and Rothwarf *et al.*² describe how they may be regulated. One critical step in activating

1. Gilbert, W. *Nature* **319**, 618 (1986).
2. Unrau, P. J. & Bartel, D. P. *Nature* **395**, 260–263 (1998).
3. Noller, H. F., Hoffarth, V. & Zimmnick, L. *Science* **256**, 1416–1419 (1992).
4. Sievers, D. & von Kiedrowski, G. *Nature* **369**, 221–224 (1994).
5. Bohler, C., Nielsen, P. E. & Orgel, L. E. *Nature* **376**, 578–581 (1995).
6. Tarasow, T. M., Tarasow, S. L. & Eaton, B. E. *Nature* **389**, 54–57 (1997).
7. Eaton, B. E. *Curr. Opin. Chem. Biol.* **1**, 10–16 (1997).
8. Hiraio, I. & Ellington, A. D. *Curr. Biol.* **5**, 1017–1022 (1995).
9. Bartel, D. P. & Szostak, J. W. *Science* **261**, 1411–1418 (1993).

NF- κ B (Fig. 1) is the rapid depletion of cytoplasmic I κ B α protein by proteolytic degradation. This is triggered by phosphorylation of I κ B α at two amino-terminal serine residues, and binding of ubiquitin then targets the phosphorylated I κ B α for degradation by the proteasome.

Several laboratories have established that signal-induced phosphorylation is accomplished by an intriguingly large IKK complex. Two IKKs — IKK- α and IKK- β — have been cloned^{3,4}, and shown to be part of multi-protein complexes in the appropriate size range (relative molecular mass, M_r , 800,000). To date, they are the only kinases known to phosphorylate I κ B α at the same residues that are modified in response to agents that activate NF- κ B (such as tumour-necrosis factor- α (TNF- α), interleukin-1 (IL-1) and lipopolysaccharide). Moreover,

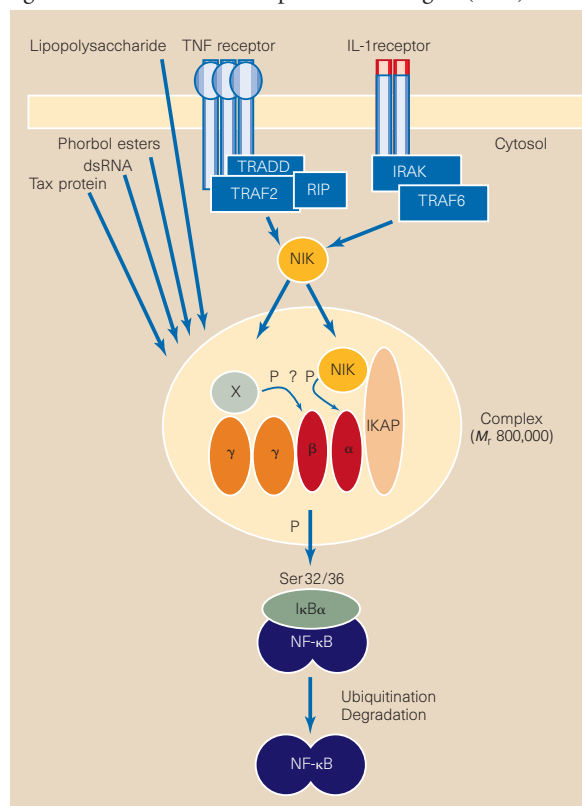


Figure 1 Putative connections between I κ B kinases, scaffold proteins and signalling molecules in the activation of NF- κ B. Tumour-necrosis factor (TNF) or interleukin-1 (IL-1) receptors initiate activation of NF- κ B by recruiting adaptor proteins (such as TRADD or TRAFs). Activated NF- κ B-inducing kinase (NIK) is released and sequestered by complexes containing I κ B kinase (IKK)- α and - β heterodimers. To be active, these kinases must associate with NF- κ B essential modifier (NEMO)/IKK- γ , which may form dimers. IKK- α and - β also interact with the IKK-complex-associated protein (IKAP). IKAP is bound by NIK, which activates IKKs. TRADD, TNF-receptor-associated death domain; TRAF, TNF-receptor-associated factor; RIP, receptor interacting protein.

IKK- α and IKK- β are both stimulated by TNF- α or IL-1.

But it is not understood how upstream signals are transmitted to the kinase complex, or whether different kinase complexes might exist to phosphorylate distinct I κ Bs. Assuming that one main kinase complex would relay diverse signals, the most interesting question is, perhaps, at which of its components the different upstream pathways converge. One possibility is scaffold proteins — these could bind IKKs and mediate their activation by assembling various upstream components. Such components may include adaptor proteins or kinases that switch on IKK activity, either by direct interaction or by phosphorylation.

Cohen *et al.*¹ and Rothwarf *et al.*² provide new clues as to how the activity of IKK- α and IKK- β is regulated. Rothwarf *et al.* have cloned a protein termed IKK- γ from purified IKK complexes. It turns out to be the human homologue of the mouse NEMO (NF- κ B essential modifier), which was isolated by Alain Israël and colleagues using a genetic approach³. These authors found that NEMO complementary DNA complemented cell lines that could not respond to the agents that normally induce NF- κ B, including TNF- α or IL-1. Thus, NEMO is essential for activation of NF- κ B by these inducers. Likewise, Rothwarf *et al.* show that expression of IKK- γ antisense DNA impairs cytokine-induced activity of IKK- α and IKK- β , degradation of I κ B and activation of NF- κ B. Furthermore, consistent with its essential role for activation of NF- κ B, NEMO/IKK- γ is found in most or all of the cellular IKK- α /IKK- β complexes in the M_r 800,000 size range^{2,5}. In the absence of NEMO, smaller complexes of M_r 300,000–400,000 are formed⁵.

Given that NEMO/IKK- γ is essential for activity of the IKK complex, is it simply a structural component, or is it involved in regulating kinase activity and signal transmission? Rothwarf *et al.* show that a carboxy-terminal domain of IKK- γ is required for TNF- α -induced — but not for basal — activity of IKK- β . This domain is also needed for stimulation of IKK activity by IL-1 or by two kinases, the NF- κ B-inducing kinase (NIK) or the mitogen-activated protein kinase/ERK kinase (MEKK1). Both NIK and MEKK1 are thought to be involved in the NF- κ B-activation pathway^{6–9}. Intriguingly, the domain contains a leucine zipper, so it could provide the interaction site for signal-transmitting factors — kinases that could activate IKK- α or IKK- β once associated, for example (Fig. 1). A likely function for NEMO/IKK- γ in the holoenzyme complex is, therefore, to physically link I κ B kinases to upstream activators.

Exactly the same function is proposed for a completely different protein cloned by Cohen *et al.*¹. The IKK-complex-associated

protein (IKAP) is associated with high-molecular-weight IKK complexes as well. Like NEMO/IKK- γ , IKAP binds to I κ B kinases directly, but it also binds to NIK. The authors found that IKAP complexes generated by co-transfection are best able to phosphorylate I κ B α when all three kinases are bound, and they are inactive when just one of the kinases is present. Furthermore, enzymatically defective IKK- β or NIK mutants inactivate the complex. A function for IKAP in stimulus-dependent I κ B kinase activity in the native kinase complex has still not been shown. But IKAP could, nonetheless, act as a scaffold protein that tethers NIK or other molecules — perhaps even NIK activators — to IKK- α and IKK- β .

One obvious question is whether IKAP and NEMO/IKK- γ are part of the same kinase complex. NEMO/IKK- γ or IKAP could both be assembly units for independent IKK complexes. Other kinases are not required for the interaction between IKK- α /IKK- β and I κ Bs, because these IKKs phosphorylate I κ B α or I κ B β directly (as shown using purified enzymes^{8,10}). But if IKAP is a scaffold protein in NF- κ B signalling, it should be part of a common complex with NEMO, because NEMO is essential for activation of NF- κ B. This is likely to be the case — all cellular, high-molecular-weight IKK- α complexes seem to contain NEMO/IKK- γ ⁵, and IKAP also seems to be bound to most cellular IKKs, and is found in similarly large complexes¹.

An intriguing observation by Cohen *et al.* is that IKAP dissociates from IKKs, but not from NIK, after stimulation by IL-1. This is in contrast to the situation with NEMO/IKK- γ , where the association with IKK- α and IKK- β is unaffected by signalling^{2,5}. Thus, IKAP could direct the assembly or disassembly of IKK complexes in response to signalling, and control deactivation of the kinase complex after stimulation. Finally, because NEMO/IKK- γ and IKAP are assembly platforms for signal transmitters, one would expect further discoveries by searching for associating factors and, perhaps, for homologues. □

Claus Scheidereit is at the Max-Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Strasse 10, 13122 Berlin, Germany.

e-mail: scheidereit@mdc-berlin.de

1. Cohen, L., Henzel, W. J. & Baeuerle, P. A. *Nature* **395**, 292–296 (1998).
2. Rothwarf, M. K., Zandi, E., Natoli, G. & Karin, M. *Nature* **395**, 297–300 (1998).
3. Maniatis, T. *Science* **278**, 818–819 (1997).
4. Stancovski, I. & Baltimore, D. *Cell* **91**, 299–302 (1997).
5. Yamaoka, S. *et al.* *Cell* **93**, 1231–1240 (1998).
6. Mercurio, F. *et al.* *Science* **278**, 860–866 (1997).
7. Malinin, N. L., Boldin, M. P., Kovalenko, A. V. & Wallach, D. *Nature* **385**, 540–544 (1997).
8. Lee, F. S., Peters, R. T., Dang, L. C. & Maniatis, T. *Proc. Natl Acad. Sci. USA* **95**, 9319–9324 (1998).
9. Karin, M. & Delhase, M. *Proc. Natl Acad. Sci. USA* **95**, 9067–9069 (1998).
10. Zandi, E. & Karin, M. *Science* **281**, 1360–1363 (1988).

Daedalus

Nuclear glue

In a cloud chamber or bubble chamber, a flying nuclear particle leaves a trail of ionized molecules; these initiate droplets or bubbles along its track. Daedalus once proposed a 'monomer chamber', in which the ions initiated polymerization of a liquid monomer. With the right chemistry, a particle should draw a single, long-chain polymer molecule along its track. An intense beam of parallel radiation might even polymerize all the monomer, giving a perfectly oriented polymeric solid.

Daedalus now has a new use for the principle. Imagine, he says, a thin film of liquid monomer between solid surfaces, traversed by flying particles. Each particle will create a single polymer chain stretching from one surface to the other. Furthermore, like the monomer, the solid surface will also be momentarily ionized by the flying particle. So it will react as well. Where it meets each surface, the polymer chain will be terminated and chemically bound to that surface.

Now a straight polymer chain is a highly improbable conformation. Its favoured shape is a random tangle. For a chain of n links, the normal end-to-end distance is not n times the length of a link, but more like $\sqrt{n}$ times. So the newly formed polymer chain is like violently stretched rubber. It will instantly contract by thermal, Brownian tangling, thus hauling the surfaces together. The next chain to form will therefore bridge a narrower gap; it will tangle in its turn, and haul the surfaces closer still. Molecules not coupled into such bridges will ultimately set solid by random polymerization.

Thus Daedalus has invented the ultimate glue. An ideal glued joint needs its surfaces butted as close as possible, and a glue that binds chemically to both of them. DREADCO's 'Radioglue' achieves both automatically. Professional engineers will be delighted. They will happily abandon their jigs, clamps and assembly rigs; Radioglue is self-tightening. Hitherto intractable metals and plastics will be glueable — nuclear radiation can ionize anything. New metal laminates, composites for airframes, high-tenacity structures, all will be Radioglued firmly together. To avoid external radiation sources, Radioglue could even be loaded with its own short-life isotope. Each decay would have to emit two particles, drawing not a straight molecule, but a dog-leg one 'hinged' at the decay site. The loss of self-tightening should be small.

David Jones

Mammary models

Clinicians screening patients for diseases such as breast cancer have to let a machine do much of the seeing for them. Theoretical modelling of the processes involved can help to ensure that reliable images are generated.

Martin Kemp

Modern diagnostic medicine involves countless life-and-death decisions taken on the basis of images formed without normal acts of seeing — using non-visible emissions, artificial perceptual systems and computerized cognition. Layered on top of artificial procedures are our own natural processes for seeing and knowing. Clinicians and technicians who use machine-generated images understand the physical means of image formation, and gain an intuitive sense of how to see what they seek, but they are unlikely to be aware of the mathematics behind image processing.

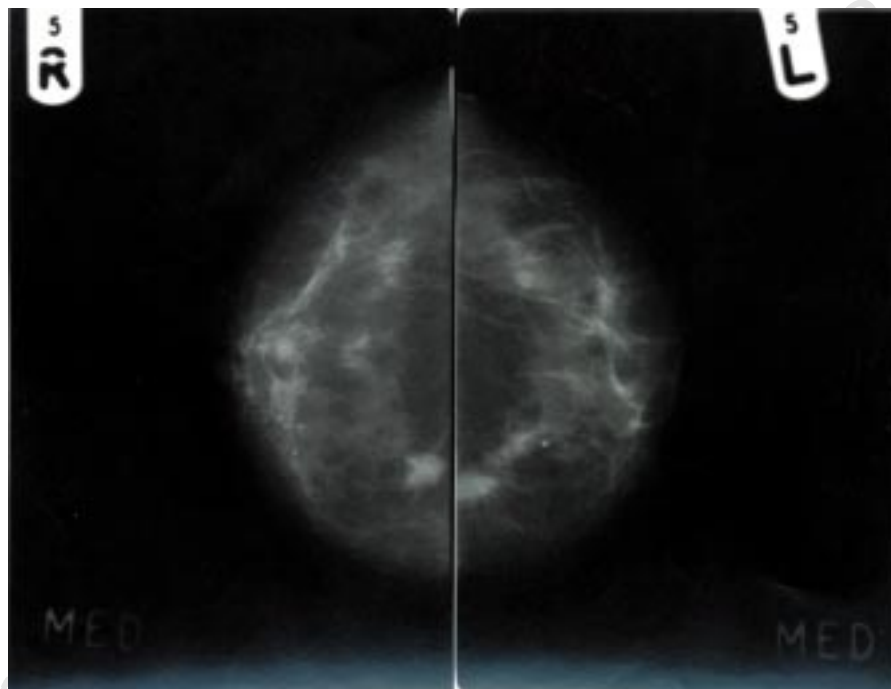
The more elusive a clear image, the more significant will be mediating procedures. Nowhere is this more apparent than in the scanning of women's breasts for tumours — an urgent matter given that breast cancer accounts for almost a fifth of cancer deaths in women.

X-ray mammography is beset by a series of difficulties, which stock imaging packages do not solve reliably. For example, an image can be sharpened in such a way that the fuzzy edges of a malignant growth acquire the defined boundary of a benign tumour.

Complex structures in breast tissue result in poor signal-to-noise ratio (always remembering that one person's noise is another person's signal). In particular, the tiny signs of crucial microcalcifications are easily lost. The low levels of 'safe' doses limit clarity. Scattering of X-ray photons blurs definition. Features at different depths in the breast are confusingly superimposed. And the compression of the breast during the procedure results in non-linear relocations of internal features.

Michael Brady and Ralph Highnam of the Department of Engineering Science at the University of Oxford are approaching these problems through the building of a highly sophisticated model of the physics of X-rays passing through breast tissue, to such good visual effect that they can closely simulate actual images and replicate the way that different imaging processes transform the information.

They can filter out the particulate noise that masks microcalcifications, augment zones of interest through 'hot light' simulation, and the definition that could theoretically be achieved with mono-energetic emissions, imitate the effect of an anti-scatter grid (which in normal practice requires a doubling of the dose), and investi-



Brady and Highnam's juxtaposed images of an actual X-ray mammogram of a tumour in the left breast (right) and a computer-modelled tumour in the right breast (left).

gate the 2D effects of 3D structures in normal and compressed configurations.

Such algorithmic modelling is a powerful tool for understanding — and a handy device for teaching. It also permits the processing of a standard image so that those features deemed to be of clinical significance can be highlighted, with a developed awareness of the parameters of what is being done. The approach is potentially applicable to any image-processing system, and is being extended at Oxford specifically to MRI breast scans, for women of any age.

Ultimately, even such meticulously controlled images have to be scrutinized by our own infinitely more complex apparatus, and judgement by the 'educated eye' of a clinically-trained connoisseur remains crucial. Indeed, it could happen that such eyes may feel uneasy working with images that look virtually identical but instinctively seem not quite the same.

Experienced and perceptive radiographers acquire a clear sense of how to work with 'their' radiologists, operating the machine to produce images that their radiologists can 'read', but may experience difficulty in seeing precisely what they want in images taken by another radiographer. This

variability does not necessarily mean that one will see a malignant tumour while another sees a benign one, or even sees nothing unusual, but it does indicate the subtly personal factors involved in the making and interpretive viewing of images that to the untrained eye reveal no useful information. Why we see certain things and overlook others remains a highly complex business.

Not only do theoretical models of different types of machine perception and cognition serve as enormously valuable aids within the technical field of diagnosis — permitting the systematic exploration of image formation and processing in a way that is not possible with actual X-rays and actual breasts — but they can also alert us to the way that the perceptual systems of individual observers work to extract meaning from the resulting radiographs.

And, beyond their medical utility, they serve to underline the selective and purposeful nature of all kinds of image processing — artificial and human — and to highlight the marvellously refined way that human vision learns to attune itself to worlds that lie quite outside the compass of our normal seen experience. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk

Mothers determine sexual preferences

The extent to which behaviour is determined by 'nurture' as opposed to 'nature' in mammals is controversial, although most recent interest has focused on genetic determinants. Here we investigate maternal influences on behavioural development by using the approach of cross-fostering between sheep and goats, which, like ourselves, form close individual attachment bonds with their offspring. We show that the emotional bond between a mother and her male offspring, rather than other social or genetic factors, may irreversibly determine these species' social and sexual preferences. Maternal influences on female offspring are weaker and totally reversible. In both sexes, visual cues from the face are important for determining attraction.

Studies of birds¹⁻³ have shown that nurture can alter the development of social and sexual preferences through parental influences, as individuals of one species cross-fostered onto another develop a preference for individuals of their cross-fostered maternal rather than genetic species. This has been called 'sexual imprinting'. However, unlike mammals, these avian species also classically imprint on the first salient animate or inanimate visual object they see, suggesting a high degree of preprogrammed inflexibility in their development of social preferences.

An important question, therefore, is whether maternal influences on social mammals that form strong attachment bonds with their offspring are similar even though such mammals exhibit increased behavioural flexibility. Maternal influences have been suggested to affect sociosexual preferences in humans, particularly males, but this has been difficult to prove. One study of three macaque monkeys cross-fostered at birth onto mothers of another macaque species showed that one monkey preferred pictures of members of its maternal rather than its genetic species, but the study made no direct assessments of social and sexual preferences⁴.

We reciprocally cross-fostered offspring between sheep (eight male and five female offspring) and goats (four male and four female offspring) at birth⁵. We tested the relative importance of the maternal bond compared with other social relationships by allowing fostered offspring social contact with members of their genetic species at all times during development. As juveniles, their play and grooming behaviour resembled that of their maternal rather than their genetic species, but species-specific patterns of aggression, climbing, feeding and vocalization were unaffected.

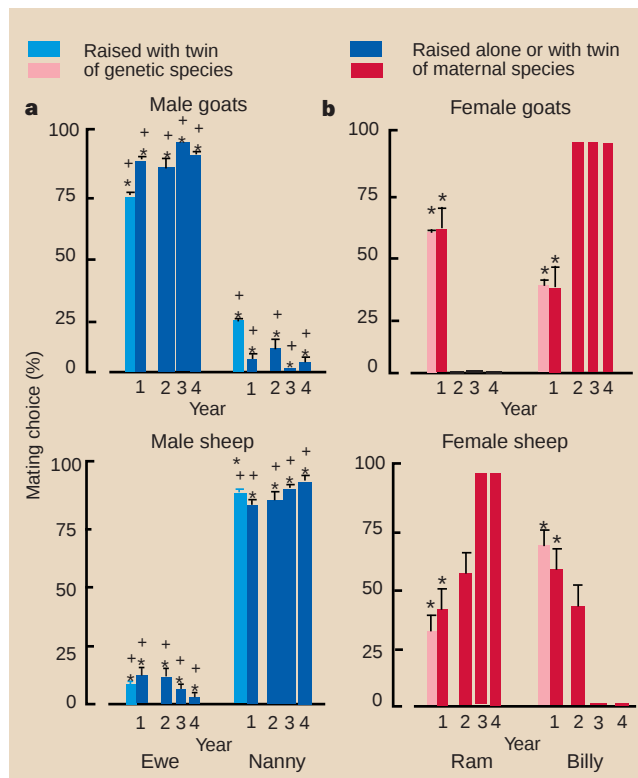


Figure 1 Maternal influence on mating choices. **a**, Percentage mating choices (mean $\pm$ s.e.m.) made by cross-fostered male sheep and goats between tethered ewes and nanny goats during 5-minute tests (30–60 tests per year). **b**, Percentage mating choices made by cross-fostered oestrus females between rams and billy goats. Normally reared sheep (9 males and 6 females) and goats (8 males and 10 females) mated exclusively with members of their genetic species. In tests in years 2, 3 and 4, animals had lived with their genetic species only after year 1. Asterisks indicate $P < 0.01$ versus normally reared animals; crosses indicate $P < 0.01$ different choices versus cross-fostered females (two-tailed Student's t -test).

During formal choice tests using adult animals, cross-fostered males strongly preferred to socialize (mean $\pm$ s.e.m. $89.1 \pm 8.3\%$ of time) and mate (Fig. 1a) with females of their maternal species. This preference was not altered even after living exclusively with their genetic species for 3 years (Fig. 1a).

In contrast, cross-fostering effects on

female social ($68.7 \pm 12.7\%$ time with maternal species females, $P < 0.001$ versus cross-fostered males; U -test) and sexual (Fig. 1b) preferences were significantly weaker and reversible within 1 to 2 years. All normally reared animals preferred social contact (males, $95.7 \pm 0.96\%$ of time; females, $94 \pm 1.1\%$ of time) and chose to mate exclusively with their genetic species.

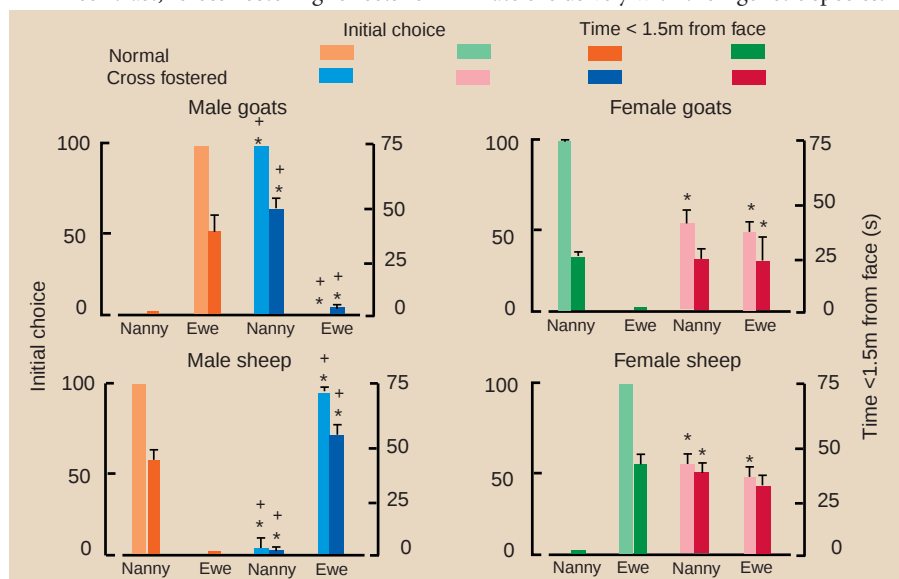


Figure 2 Mean $\pm$ s.e.m. percentage initial choice of nanny goat or ewe faces made by normal and cross-fostered male and female sheep and goats ($n=5$ normal and $n=4$ cross-fostered animals for each species and sex) and durations spent within 1.5 metres of the faces during 120-second tests. Asterisks indicate $P < 0.05$ versus normally raised animals; crosses indicate $P < 0.05$ compared with proportion of choice/duration of time of nanny versus ewe face in females (two-tailed Student's t -test).

We also determined whether sibling bonds might reduce the impact of the maternal bond, as lambs and kids form close bonds with a twin. However, cross-fostering opposite-sex twins of the same genetic species (kids, $n=10$; lambs, $n=8$) did not prevent the maternal influence on preferences from occurring (Fig. 1a, b).

Sheep, like primates, can recognize individuals using facial cues^{6,7}. In choice tests using pictures of sheep and goat faces, we found that these alone could elicit preference for females of the maternal species and that effects were again stronger in males (Fig. 2). Thus the face appears to be an important source of attraction.

This strong maternal influence on social and sexual preferences may function to prevent cross-species matings. However, it has been argued for avian species that sexual imprinting may also ensure an optimal outbreeding strategy, as cross-fostered individuals prefer mates that differ only slightly in appearance from their mothers⁸. The fact that male offspring are affected more than females, and apparently for life, is evidence that they are indeed more potently influenced by their mothers. This indirectly supports Freud's concept of the Oedipus complex and suggests that males may also be less able than females to adapt to altered social priorities.

Keith M. Kendrick, Michael R. Hinton, Khia Atkins

Laboratory of Cognitive and Developmental Neuroscience, The Babraham Institute, Babraham, Cambridge CB2 4AT, UK
e-mail: keith.kendrick@bbsrc.ac.uk

Martin A. Haupt, John D. Skinner

Mammal Research Institute, University of Pretoria, Pretoria 0002, South Africa

1. Lorenz, K. J. *Ornithol.* **83**, 137–213 & 289–413 (1935).
2. Immelmann, K. Z. *Tierpsychol.* **26**, 677–691 (1969).
3. ten Cate, C. in *Perspectives in Ethology* Vol. 8 (eds Bateson, P. P. G. & Klopfer, P. H.) 243–269 (Cambridge Univ. Press, New York, 1989).
4. Fujita, K. *Primates* **34**, 141–150 (1993).
5. Kendrick, K. M., da Costa, A. P., Hinton, M. R. & Keverne, E. B. *Appl. Anim. Behav. Sci.* **34**, 345–357 (1992).
6. Kendrick, K. M., Atkins, K., Hinton, M. R., Heavens, P. & Keverne, E. B. *Behav. Proc.* **38**, 19–35 (1996).
7. Kendrick, K. M. & Baldwin, B. A. *Science* **236**, 448–450 (1987).
8. Bateson, P. *Nature* **273**, 659–660 (1978).

All limbs are not the same

Recent papers published in *Nature* have assumed that the mechanism of limb development in all vertebrates is the same^{1,2}. But if mechanisms were conserved in all tetrapods, we should expect to find them in amphibians as well as in amniotes. We have examined the expression of eight important signalling and regulatory molecules in *Xenopus* limb development and

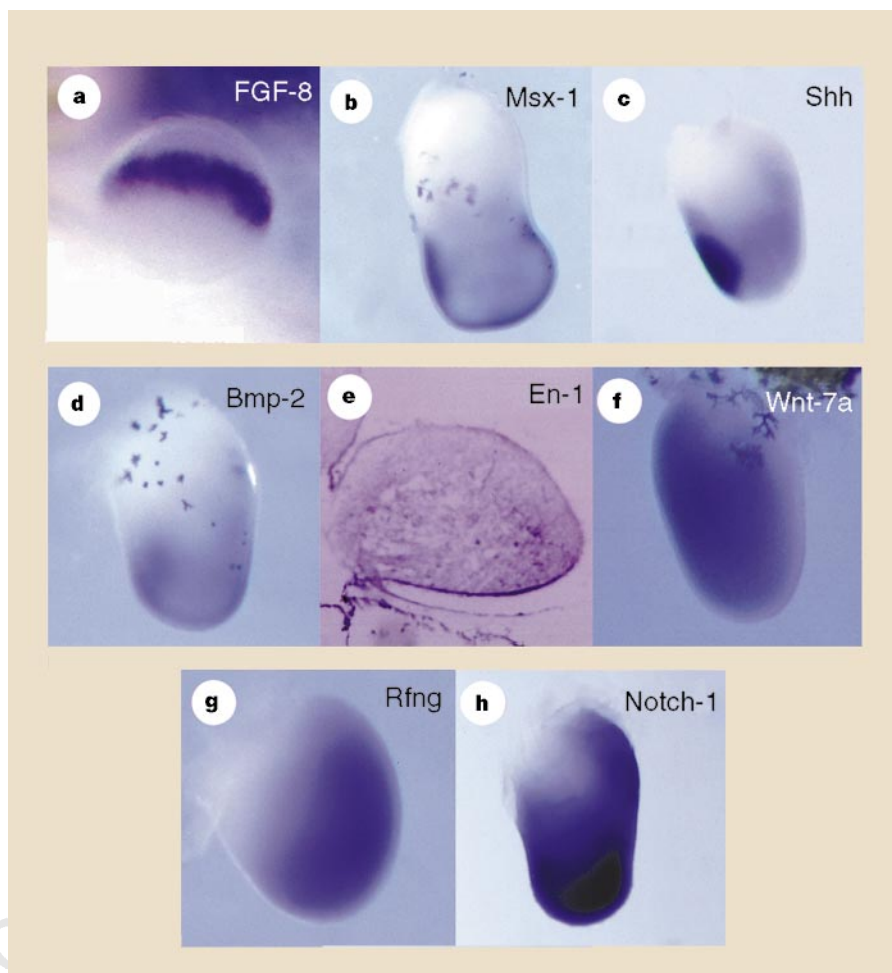


Figure 1 *In situ* hybridization in *Xenopus* limb buds at stage 50/51 (a, c–h) or stage 53 (b) for the eight genes studied.

find that the assumption is not correct. This finding has obvious implications for our understanding of limb development and evolution.

The prevailing view of limb patterning, based on experiments in chick and mouse, involves three distinct signalling centres, each controlling the differentiation of structures along one of the anatomical axes of the limb bud: proximodistal, anteroposterior and dorsoventral³. By using various reagents we found that the proximodistal and anteroposterior systems in *Xenopus* appear similar to the amniote species, whereas the dorsoventral system appears to be different (Fig. 1).

The model for dorsoventral patterning³ involves activation of the transcription factor En-1 in the ventral ectoderm at an early stage. Expression of En-1 represses the expression of two signalling molecules, Wnt-7A and Radical fringe (Rfng), which are therefore made only in the dorsal ectoderm. The Wnt-7A signal causes the dorsal mesenchyme to form dorsal structures. The Rfng signal participates in the induction of the apical ectodermal ridge (AER), probably by potentiating the action of another signal, Serrate, on its receptor Notch-1 (refs

4, 5). We have examined the expression of the genes *en-1*, *Wnt-7A*, *Rfng* and *Notch-1* in *Xenopus* limb buds. Of these, only *en-1* is expressed in the expected position, the ventral epidermis. The other three do not show the expected regionalization, but are expressed in a diffuse manner throughout the limb bud in both ectoderm and mesenchyme. We have confirmed that they really are expressed, and that the diffuse staining is not just nonspecific background, by RNase protections (Fig. 2).

The proximodistal pattern of amniote limbs arises from the sequential formation of structures from a mesenchymal progress zone, the developmental lability of which is maintained by fibroblast growth factors (FGFs) secreted by the AER³. In *Xenopus* there is an apical band of expression of FGF-8, which presumably functions as the AER. There is also expression of the transcription factor Msx-1 in the underlying progress zone. The anteroposterior pattern arises in response to the secretion of Sonic hedgehog (Shh) from the zone of polarizing activity on the posterior side of the mesenchyme⁶. *Xenopus* has a similar localized expression of Shh, and a similar expression of Bmp-2, in both the zone of polarizing

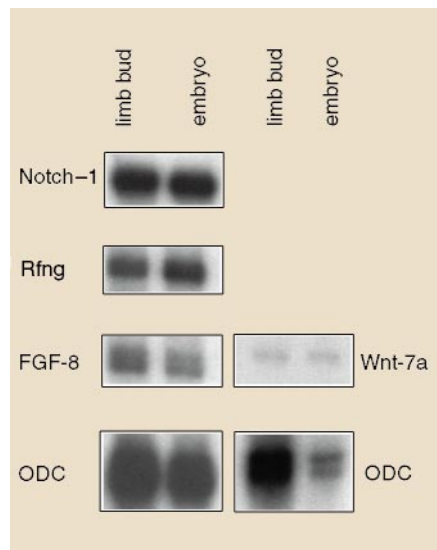


Figure 2 RNase protection analysis of stage 50/52 *Xenopus* limb buds showing the three diffusely expressed gene products together with FGF-8. Results for the same genes in stage 33 embryos are shown as positive controls. Ornithine decarboxylase (ODC) is a loading control. Detailed methods are available from the authors.

activity and the AER. These observations suggest that the mechanisms in proximodistal and anteroposterior axes are indeed conserved.

Amphibia was the first tetrapod class to diverge from the vertebrate lineage, so our results suggest that the proximodistal and anteroposterior patterning systems have indeed been conserved across the tetrapods. The same cannot be said of the dorsoventral system, however. Our results show that *En-1* is not repressing the dorsal genes, and that it is not necessary to have restricted expression of *Wnt-7A*, *Rfng* or *Notch-1* to make a limb. Because all these genes are expressed diffusely, it remains possible that there is a permissive requirement for their products, but a mechanism depending on interaction across boundaries^{4,5} does not seem likely.

Some differences between amphibian and amniote limb development that might account for a difference in the mechanism of dorsoventral patterning are known⁷, including a more complex muscle pattern in amniotes and the regenerative ability shown by many amphibians, including *Xenopus* in the larval stage^{8,9}. We have examined the expression of the same eight genes in regeneration blastemas and find very similar results to the original limb buds (data not shown). Whatever the reason for the difference, our results show that the accepted mechanism for dorsoventral patterning is not a necessary feature of limb development but is, at its most general, a derived feature shown by the amniotes.

Bea Christen, Jonathan M. W. Slack
Developmental Biology Programme,
Department of Biology and Biochemistry,

University of Bath,
Bath BA2 7AY, UK.
e-mail: bssbc@bath.ac.uk

1. Bushdid, P. B., et al. *Nature* **392**, 615–618 (1998).
2. Kanegae, Y., Tavares, A. T., Izpisua Belmonte, J. C. & Verma, I. M. *Nature* **392**, 611–614 (1998).
3. Johnson, R. & Tabin, C. J. *Cell* **90**, 979–990 (1997).
4. Laufer, E. et al. *Nature* **386**, 366–373 (1997).
5. Rodriguez-Esteban, C. et al. *Nature* **386**, 360–366 (1997).
6. Riddle, R. D., Johnson, R. L., Laufer, E. & Tabin, C. *Cell* **75**, 1401–1416 (1993).
7. Hinchliffe, J. R. & Johnson, D. R. *The Development of the Vertebrate Limb* (Oxford Univ. Press, 1980).
8. Dent, J. N. J. *Morphol.* **110**, 61–78 (1962).
9. Maden, M. J. *Embryol. Exp. Morphol.* **63**, 243–265 (1981).

Does vitamin C have a pro-oxidant effect?

The conclusions of the Scientific Correspondence by Podmore *et al.*¹ are undermined by some unaddressed scientific issues. Podmore *et al.* gave vitamin C supplements of 500 milligrams to healthy volunteers, and measured levels of vitamin C in plasma and of oxidative DNA adducts in lymphocytes before, during and after supplementation.

Because the reported changes in levels of 8-oxoguanine and 8-oxoadenine occurred in lymphocytes, the relevant vitamin C concentrations are not those in plasma but those in lymphocytes; however, these data were not presented. Because millimolar concentrations of ascorbate are found in lymphocytes, measurements of vitamin C are practical using as few as 0.25×10^6 cells and a sensitive electrochemical high-performance liquid chromatography assay^{2–4}. Available cell numbers should not have been a limiting problem because few cells are needed for the measurement compared with the number that can be isolated from peripheral blood^{2,3,5}.

Although the data on plasma vitamin C concentrations presented by Podmore *et al.*¹ were expressed as percentage changes only, true intracellular vitamin C concentrations can still be calculated. Because the vitamin C supplement produced a 60% increase in plasma concentrations of vitamin C, and because 500 mg is a saturating dose of vitamin C, the initial plasma vitamin C concentration was roughly 50 μ M (refs 5,6). However, at this concentration lymphocytes are already saturated, with an intracellular ascorbate concentration of approximately 3 mM (refs 5,6). Increasing the extracellular plasma concentration of ascorbate above 50 μ M will not affect the intracellular concentration further.

If the reported changes in oxidative DNA adducts could not be due to changes in intracellular lymphocyte vitamin C concentrations, what is responsible? The concentrations of 8-oxoguanine reported by Podmore *et al.* are 25–120 times more

than those reported by others^{7–9}. These measurements are notoriously difficult to make and are easily increased artefactually by oxidation⁹. The possibility of oxidation is increased further if mononuclear cells rather than pure lymphocytes were isolated, because activated contaminating monocytes can produce superoxide and other oxidants. The purity of the isolated lymphocytes was not indicated by Podmore *et al.* In addition, without proper precautions ascorbate can act as a pro-oxidant. By increasing plasma ascorbate levels with supplements, the authors may have increased the probability of an oxidation artefact occurring.

Because Podmore *et al.* used an experimental design that did not take into account the fact that cells become saturated with vitamin C at lower doses than does plasma, an inappropriate choice of tissue sampled for vitamin C, unclear lymphocyte-isolation procedures, and assays that may be biased by oxidation artefacts, we believe that their conclusions are not justified by the data.

Mark Levine, Rushad C. Daruwala, Jae B. Park, Steven C Rumsey, Yaohui Wang
Molecular and Clinical Nutrition Section,
Building 10, Room 4D 52, MSC 1372,
National Institute of Diabetes and Digestive
and Kidney Diseases, National Institutes of Health,
Bethesda, Maryland 20892-1372, USA

1. Podmore, I. D. et al. *Nature* **392**, 559 (1998).
2. Washko, P., Rotrosen, D. & Levine, M. *FEBS Lett.* **260**, 101–104 (1990).
3. Bergsten, P. et al. *J. Biol. Chem.* **265**, 2584–2587 (1990).
4. Washko, P. W., Welch, R. W., Dhariwal, K. R., Wang, Y. & Levine, M. *Anal. Biochem.* **204**, 1–14 (1992).
5. Levine, M. et al. *Proc. Natl Acad. Sci. USA* **93**, 3704–3709 (1996).
6. Graumlich, J. F. et al. *Pharm. Res.* **14**, 1133–1139 (1997).
7. Nakae, D., Mizumoto, Y., Kobayashi, E., Noguchi, O. & Konishi, Y. *Cancer Lett.* **97**, 233–239 (1995).
8. Nakajima, M., Takeuchi, T. & Morimoto, K. *Carcinogenesis* **17**, 787–791 (1996).
9. Helbock, H. J. et al. *Proc. Natl Acad. Sci. USA* **95**, 288–293 (1998).

Podmore *et al.*¹ reported a decrease in 8-oxo-7,8-dihydroguanine and an increase in 8-oxo-7,8-dihydroadenine in lymphocyte DNA after intake of 500 milligrams of vitamin C daily for 6 weeks. In our view, four issues compromise the validity of their findings.

First, there are severe problems resulting from increased oxidation during the DNA isolation and extraction necessary^{2,3} for gas chromatography mass spectrometry. Podmore *et al.* reported high values in lymphocyte DNA of 30 molecules of 8-oxo-7,8-dihydroguanine per 100,000 guanines, but did not acknowledge these problems or refer to related published studies (see, for example, ref. 4). Artefactual oxidation *ex vivo* in their study cannot be excluded and hence their results are difficult to interpret.

Second, the study design is without randomization or true placebo control, nor

is it double-blinded; thus, time and other effects unrelated to those of vitamin C cannot be excluded. Third, Podmore *et al.* did not mention whether any of their subjects smoked. Fourth, the authors did not reference relevant previous studies.

We have reported a 59-fold variation in guinea-pig liver vitamin C concentrations as a result of dietary manipulation, without any change in oxidized guanine levels in the liver³, for example. In a two-month randomized, placebo-controlled trial in which we gave 38 smoking men 500 mg vitamin C every day, we found no change in oxidative DNA damage measured by 24-hour urinary excretion of 8-oxo-7,8-dihydro-2'-deoxyguanosine, as compared with 19 smoking men who received a placebo⁶. Urinary excretion of 8-oxo-7,8-dihydro-2'-deoxyguanosine is interpreted as a total-body average measurement of the rate of DNA oxidation, a theory supported by recovery of 2-nitro-propane-induced excess organ 8-oxo-7,8-dihydro-2'-deoxyguanosine in excreted urine in animals⁷.

Measurements of the oxidized bases 8-oxo-7,8-dihydroguanine and 8-oxo-7,8-dihydroadenine, or of the corresponding nucleosides 8-oxo-7,8-dihydro-2'-deoxyguanosine and 8-oxo-7,8-dihydro-2'-deoxyadenosine, in lymphocytes represents concentrations of oxidized bases in those cells only and their relevance to other organs, for example target organs, is not known. Neither is it known whether one or both of the two measurements is relevant for risk assessment.

We think it is too soon to say whether supplemental doses of vitamin C exert pro-oxidant or mutagenic effects. Further well-designed trials with an analysis that would eliminate artificial *ex vivo* oxidation are needed to resolve this question.

Henrik E. Poulsen, Allan Weimann

Department of Clinical Pharmacology Q 7642,
Rigshospitalet, University of Copenhagen,
DK-2200 Copenhagen N, Denmark
e-mail: henrikep@rh.dk

Jukka T. Salonen, Kristiina Nyyssönen

Research Institute of Public Health,
University of Kuopio, Kuopio, FIN-70211, Finland
Steffen Loft

Institute of Public Health,
University of Copenhagen, Panum 18-5,
DK-2100 Copenhagen, Denmark

Jean Cadet, Thierry Douki,

Jean-Luc Ravanat

DRFMC/SCIB, Laboratoire Lésions des Acides
Nucléiques, CEA/Grenoble, F-38054 Grenoble,
Cedex 9, France

6. Prieme, H., Loft, S., Nyyssönen, K., Salonen, J. T. & Poulsen, H. E. *Am. J. Clin. Nutr.* **65**, 503–507 (1997).

7. Deng, X.-S., Tuo, J., Poulsen, H. E. & Loft, S. *Free Rad. Res.* **28**, 323–333 (1998).

Podmore et al. reply — In our Scientific Correspondence¹ we reported changes in lymphocyte DNA to two established 'markers' of oxidative DNA damage after daily supplementation of healthy volunteers with 500 mg vitamin C for six weeks. One marker, 8-oxoadenine, increased, whereas the other, 8-oxoguanine, decreased. On this basis, vitamin C seems to act as both a pro-oxidant and an antioxidant *in vivo*. Our contribution has stimulated many comments, most of which were addressed in our original manuscript to *Nature*, which was shortened for publication at the editors' request. The full details are available from us on request.

The suggestion made above by Poulsen *et al.* and Levine *et al.* that artefactual oxidation has occurred during lymphocyte isolation and DNA extraction is based entirely on the measurement of a single marker, 8-oxoguanine. As founder members of the European Standards Committee on Oxidative DNA Damage (ESCODD), we are aware of the discrepancies between 8-oxoguanine measurement by different techniques. Indeed, with this in mind we have measured 8-oxoguanine in DNA samples from healthy volunteers by two additional assays using high-performance liquid chromatography: one measuring the base, and the other, widely used, its 2'-deoxynucleoside derivative, 8-oxo-2'-deoxyguanosine (Cooke, M. S. *et al.*, manuscript in preparation). All three assays show the same decrease in 8-oxoguanine on vitamin C supplementation.

We disagree with Poulsen *et al.* about the need for randomization or a double-blind study, as we were not reporting a clinical trial requiring a subjective interpretation of efficacy. The design of our study included a correct placebo treatment. Methods for improving DNA extraction are currently evolving through ESCODD. The important question remains of why, if 8-oxoguanine and 8-oxoadenine are generated solely as artefacts, one increases and the other decreases on supplementation but both return to baseline/placebo values following washout.

We do not consider the references cited by Poulsen *et al.* to be relevant to our study on human volunteers (all of whom were non-smokers, as smoking is known to reduce levels of plasma ascorbate). One citation refers to a study measuring 8-oxoguanine in guinea-pig liver on "dietary manipulation"². The other used only measures of 8-oxo-2'-deoxyguanosine in urine³, which has not yet been shown to derive directly from DNA. However, its relevance as a marker of the total-body average rate of

DNA oxidation cannot be ruled out. Using a monoclonal antibody to detect 8-oxo-2'-deoxyguanosine, we observe an increase in urine following supplementation of the volunteers with vitamin C, a finding that corresponds to loss of 8-oxoguanine from DNA (Cooke, M. S. *et al.*, manuscript in preparation).

The mean plasma ascorbate concentration for 30 healthy volunteers (14 males and 16 females, aged between 17 and 49) in our study was 51.3 µM, increasing to 80.4 µM on vitamin C supplementation (Cooke, M. S. *et al.*, manuscript in preparation). Levine *et al.* argue that lymphocytes saturate when plasma ascorbate levels reach 50 µM. We do not find their argument compelling as it seems to be based on a depletion-repletion study of seven healthy volunteers, all males aged between 20 and 26 (ref. 4). In addition, we find considerable individual variation of plasma ascorbate with presupplementation values as low as 8 µM, equivalent to an intracellular ascorbate concentration of approximately 1.5 mM (less than 50% of the stated saturation value)⁴. It is therefore reasonable to conclude that saturation of plasma with vitamin C does not occur with at least 50% of our volunteers before supplementation.

Furthermore, we observe highly significant positive and negative correlations between plasma ascorbate concentration and levels of 8-oxoadenine and 8-oxoguanine, respectively, which we believe suggests that vitamin C is influencing levels of both markers in a divergent manner (Cooke, M. S. *et al.*, manuscript in preparation). This differential effect, which cannot be explained as methodological artefact, is related to the possible induction of specific repair of 8-oxoguanine. There is a precedent for redox regulation of repair of this lesion⁵.

In conclusion, our results show a definite increase in 8-oxoadenine after supplementation with vitamin C. This lesion is at least ten times less mutagenic than 8-oxoguanine⁶, and hence our study shows an overall profound protective effect of this vitamin.

Ian D. Podmore, Helen R. Griffiths,

Karl E. Herbert, Nalini Mistry,

Pratibha Mistry, Joseph Lunec

Division of Chemical Pathology,
Centre for Mechanisms of Human Toxicity,
University of Leicester,
Leicester LE1 9HN, UK
e-mail: idp@le.ac.uk

1. Podmore, I. D. *et al. Nature* **392**, 559 (1998).

2. Cadenas, S., Barja, G., Poulsen, H. E. & Loft, S. *Carcinogenesis* **18**, 1833–1836 (1997).

3. Prieme, H., Loft, S., Nyyssönen, K., Salonen, J. T. & Poulsen, H. E. *Am. J. Clin. Nutr.* **65**, 503–507 (1997).

4. Levine, M. *et al. Proc. Natl Acad. Sci. USA* **93**, 3704–3709 (1996).

5. Lee, H.-S. *et al. Free Rad. Biol. Med.* **24**, 1193–1201 (1998).

6. Wood, M.L. *et al. Nucleic Acids Res.* **20**, 6023–6032 (1992).

1. Podmore, I. D. *et al. Nature* **392**, 559 (1998).

2. Cadet, J., Douki, T. & Ravanat, J.-L. *Environ. Health Perspect.* **105**, 1034–1039 (1997).

3. Helbrook, H. J. *et al. Proc. Natl Acad. Sci. USA* **95**, 288 (1998).

4. Collins, A., Cadet, J., Epe, B. & Gedik, C. *Carcinogenesis* **18**, 1833–1836 (1997).

5. Cadenas, S., Barja, G., Poulsen, H. E. & Loft, S. *Carcinogenesis* **18**, 2373 (1997).

The cowboys in Indian science

Nonsense in Indian Science

by Dilip M. Salwi

Konark Publishers, A-149 Main Vikas Marg,
Delhi 110 092; 1998. 124pp. Rs125

Pushpa M. Bhargava

Few Indian scientists would disagree with the overall perception of the country's science as it emerges from Dilip Salwi's book: that science in India today is characterized by a pervasive mediocrity; lack of originality and of the courage to challenge absurdity and defend truth; widespread plagiarism and the denial of credit where it is due; fear of excellence; complacency; self-glorification; cronyism of every description; careerism without commitment; lack of professional, social or financial accountability; ignorance of scientific methods and the qualities of knowledge gained through their use; lack of the realization that science has a social responsibility and is increasingly forging and strengthening links with other areas of human endeavour; fraud, dishonesty and lack of integrity; sycophancy and crisis of leadership; the scientist-politician syndrome that drives even able and intelligent scientists to covet bureaucratic or political power; and false, premature and unsubstantiated claims. (The list is not complete.)

All this probably exists elsewhere too, but nowhere is it on the same scale and so endemic as in India. What makes it fascinating is that all this mediocrity and dubious behaviour coexists — not always peacefully — with a significant amount of excellence in almost every field.

A vast majority of Indian scientists — especially those occupying high positions — believe in miracles, homeopathy and astrology; to subscribe to them can be a prerequisite of power, awards and rewards.

Salwi is right when he says that our system of authorship of scientific papers is, by and large, ridiculous; that, in our scientific hierarchy, "power is knowledge"; that nuisance value and not merit counts when it comes to obtaining gains such as promotions; that the emphasis is on equipment and the empire that one has built, not on work or productivity; that our quality of journals is abysmally poor; that our scientific managers hardly ever read serious science; and that there is an all-round resistance to people working on truly new and frontier areas.

Our systems for assessing merit at times border on the ridiculous. It would not be difficult to find fellows of our national academies who, when they were elected (at the age of 45 or above), did not have a single independent paper representing their work in a journal indexed in *Current Contents*. Salwi is right on all these counts.



Lonely centre of excellence: the Tata Institute of Fundamental Research in Bombay.

But the book is not without faults. First, several statements made are not true. For example, Salwi says "the country is unable to establish centres of excellence or relevance". This is to ignore the Tata Institute of Fundamental Research in Bombay, the Centre for Cellular and Molecular Biology in Hyderabad, the Centre for Development of Telecommunications in Delhi, and the Centre for Development of Advanced Computing in Pune. We can establish centres of excellence, but we rarely maintain them, for in our country excellence becomes its own enemy.

Salwi also mistakenly claims that the work on biotechnology and on AIDS in India has been done with foreign money. In fact, despite enormous sums from various national sources we have done very little in these areas anyway. While Salwi's generalizations are mostly sound, he gives no examples to back them up.

Then there are significant individual exceptions to everything he says. The life and work of some of those outstanding people who helped to build modern India still hold important and relevant lessons for the future of science here. Salwi does not touch on the reasons for the nonsense in Indian science. He makes no comment on the historical background against which science evolved in India after Independence, nor does he suggest any ways to improve the situation.

The problems currently besetting Indian science can be broadly attributed to a failure to democratize knowledge after Independence, even after its establishment in the political realm. Meaningful and effective education in India today is the prerogative of less than 3 per cent of the population. Our scientific and technical manpower of 2.5 million is derived from a real take-off population of less than 30 million. The possible talents and contributions of the vast majority of the population remain untapped.

But a lack of democratization also obtains

within the science that does take place. After independence, the government failed to exercise sufficient care and objectivity in appointing scientists to senior positions. After two decades, a scientific mafia had begun to emerge, demanding sycophancy in exchange for support, and creating wholly the wrong environment for the scientific traditions of questioning and experimentation to grow. The excellence that survived became fragmented, and could not act collectively.

To improve the situation, the Indian government must recognize the existence of this scientific mafia and the damage it has done to Indian science, and the importance of democratizing education at all levels in a meaningful and productive way. Secondly, the government must completely overhaul the system for senior appointments: the criteria used must be made public, and those who recommend appointments must be made accountable for the success or failure of the people recommended.

Thirdly, the government must seek accountability from scientists. It must carry out exemplary investigations and reviews of scientific organizations, using stringent international criteria and reviewers. One such review must be of the three science academies — the India National Science Academy, the India Academy of Sciences, and the National Academy of Sciences — which have accomplished virtually nothing and have zero public credibility in spite of substantial government expenditure on them. I recall the many congratulatory messages I had from scientists and others when I resigned from the fellowship of these academies in 1994.

Government alone cannot remedy the situation. Indeed, every Indian government since Independence, and the Indian public, though at times ignorant, have been extremely supportive of science. The most important conclusion that emerges from

Salwi's book is that the responsibility for this parlous state lies largely with scientists themselves. Our scientists with conscience must show courage, strength of conviction, concern and commitment. Those who uphold excellence must act in concert to throw out mediocrity and corruption — intellectual, moral and financial. This is, perhaps, beginning to happen and, therefore, there is hope.

Nonsense in Indian Science makes somewhat laborious reading and often leaves the reader hanging. It is not a satirical or entertaining book, as the author claims; the punch-lines are rarely funny. On balance, however, it is an honest and brave work, deserving the notice of those who mismanage our field. □

Pushpa M. Bhargava is in the Centre for Cellular and Molecular Biology, Anveshna Consultancy, 12-13-414/4, Street No 1, Tarnaka, Hyderabad - 500 017, India.

Bodies of evidence

On the Fabric of the Human Body.
Book One: The Bones and Cartilages
by Andreas Vesalius. Translated by William Richardson and John Carman
Norman Publishing, 720 Market Street, San Francisco, CA 94102-2502: 1998. 416pp. \$225

The Complete Visible Human
by Heinz-Otto Peitgen, Wilhelm Berghorn and Matthias Biel
Springer Verlag: 1998. Two CDs. \$129, £120.88

Bernard Wood

The waiting room was dowdy and crowded. Most of its occupants showed varying degrees of bewilderment and apprehension. Some were evidently resigned, others frankly belligerent. A man came in, called my name, led me into the courtroom, pointed to a chair, but indicated that I should remain standing. Three magistrates looked at me expectantly.

I introduced myself as the newly appointed professor of anatomy at the local medical school, and that in order to carry out my duties I needed to become a 'licensed teacher of anatomy'. It was the magistrates' turn to look bewildered.

I explained that, because of the unacceptable activities of body-snatchers such as Burke and Hare, the then government had presented to Parliament the first of several Anatomy Acts to regulate the circumstances under which human bodies could be held for dissection. I went on to explain that responsibility for ensuring compliance with the act rests with Her Majesty's Inspector of Anatomy, who reports to the Queen's Privy Council, to whom I had to apply for my licence. Finally, I told them that all applications to be a licensed teacher had to be countersigned by three magistrates. The magistrates con-

ferred, examined my application, signed it and it was duly passed back to me. I murmured my thanks for their time and attention, and with obvious relief they turned to more familiar matters, such as traffic violations, alcohol- and drug-related offences and unpaid fines.

These arrangements are a peculiarly British, but remarkably effective, way of regulating human dissection. Across the world societies recognize that human dissection is necessary in order to train those who deal with the sick. Without a sound knowledge of normal anatomy, how can practitioners ensure that they can recognize abnormality? Without an understanding of the nervous system, how can you appreciate that the pathways and connections of sensory nerves are such that pain in one location may mean disease in another? Likewise, programmes of development do not always go to plan. Organs may end up in the 'wrong' place, joints that

should be made are not made, and natural drainage tubes do not always meet up with the organ to be drained. Accurate information such as this needs to be communicated to students of medicine and the other health care professions so that patients' physical signs and symptoms can be interpreted and converted into diagnoses.

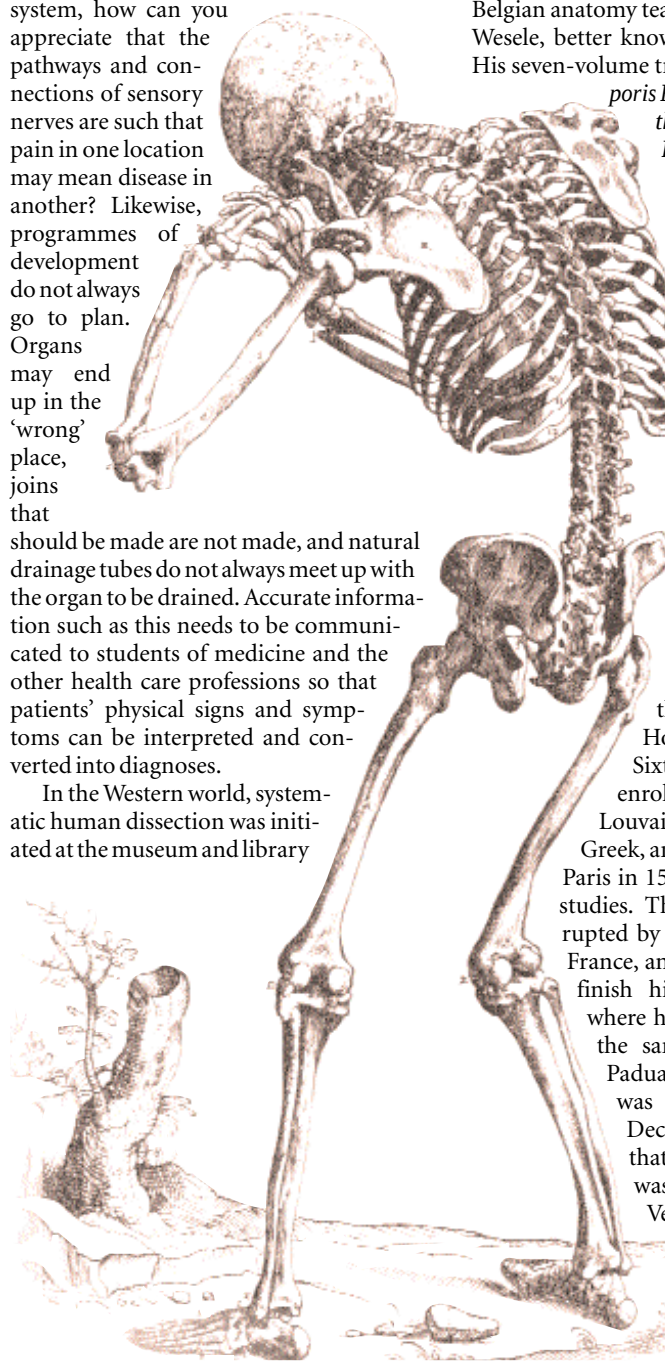
In the Western world, systematic human dissection was initiated at the museum and library

in Alexandria, and some of the first recorded dissections were carried out by Herophilus and Erasistratus in the third century BC. Much was achieved by these pioneers, but the tradition of gaining firsthand evidence of human morphology by dissection did not last long. By Galen's time, in the second century AD, the practice had effectively ceased. Thus, the teachers of anatomy were no longer accumulating any new knowledge, nor were they able to verify the information passed down to them by their predecessors. For close to a millennium and a half, human anatomy teaching consisted of the transmission of dogma.

But all this was to be changed by a young Belgian anatomy teacher called Andreas van Wesele, better known as Andreas Vesalius. His seven-volume treatise, *De Humani Corporis Fabrica Libri Septem* (*On the Fabric of the Human Body*), must rank among the foremost contributions to biology. It was much more than just the sixteenth-century equivalent of Gray's *Anatomy*: it reintroduced and secured the philosophy that the fruits of investigation should take the place of speculation and dogma as the basis of our knowledge of human structure and function.

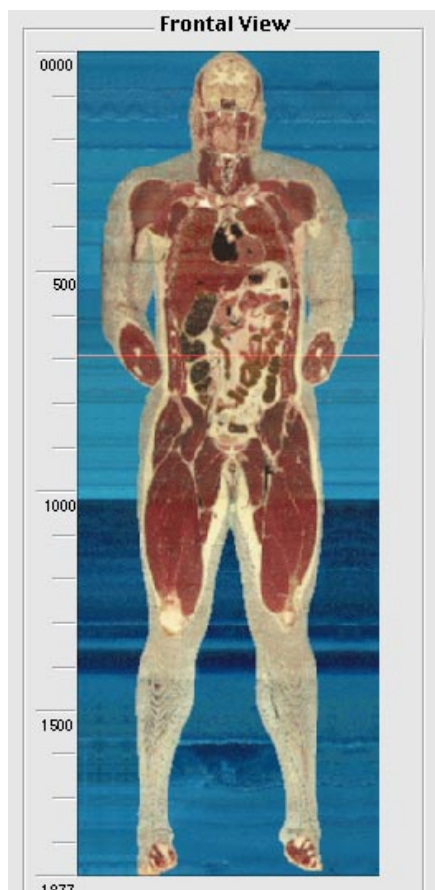
Vesalius was born in Brussels on the first day of 1514, the son of the chief apothecary to the court of Charles V, Holy Roman Emperor. Sixteen years later, he enrolled in the University of Louvain to study Latin and Greek, and then he transferred to Paris in 1533 to begin his medical studies. These were rudely interrupted by Charles V's invasion of France, and Vesalius was forced to finish his studies in Louvain, where he graduated in 1537. In the same year he moved to Padua. It is recorded that he was granted an MD on 5 December 1537. Later in that year, and still only 23, he was appointed by the Venetian Senate to teach anatomy and surgery.

In the preface to the *Fabrica* he chronicles his dissatisfaction with his anatomy teaching, and, while still in Paris, his skills in dissection, coupled with his frustration at his



Boning up: Vesalius's meticulous research overturned existing beliefs.

teaching, and, while still in Paris, his skills in dissection, coupled with his frustration at his



Slice of life: one of more than 7,000 views of the human body in *The Complete Visible Human*.

incompetent teachers, led him to take over the dissection duties. Apparently, dissections were routinely performed by assistants while the professor intoned from a textbook at a safe distance from the cadaver, usually without making any direct reference to it. Students were regularly told one thing, and glimpsed another.

Vesalius clearly had a talent for teaching, and by 1538 he had published, in Venice, six of his teaching charts as *Anatomical Tables*. During a visit to Bologna in 1540, when Vesalius was given the opportunity to compare the skeletons of a monkey and a human, he realized that many of the errors in the Galenic texts had been made because the information they contained had been based on monkey, and not human, anatomy. It was this realization that set the tone for the *Fabrica*. That Vesalius was able to perform all the dissections, record the descriptions and sketch the drafts of the illustrations for his *magnum opus* is astonishing. That all this was achieved in time for the *Fabrica* to be published in 1543, when Vesalius was still only 29, is incredible. The scale and rate of progress of his achievement is without parallel in biology.

Until now, Vesalius has not been well served by translators. Some lacked linguistic competence, others anatomical expertise. But, in what must rank as one of the publish-

ing and scientific literary achievements of the decade, the classicist William Richardson, in collaboration with anatomist John Carman, have produced a quite stunning translation of the first book of the *Fabrica*, *The Bones and Cartilages*.

They have worked hard not to unconsciously modernize the technical language. Although the anatomical terms they use reflect Vesalius's sometimes antiquated choice, they faithfully reproduce the style used in the original. The language is direct, yet preserves the complexity and style of Latin as used in the mid-sixteenth century. It is to Richardson's credit that he has brought the text to life, and one can almost hear Vesalius speaking as one reads the iconoclastic preface. The plates are believed to have been produced in Titian's studio, and they are handsomely reproduced. The publishers are to be congratulated on the presentation and quality of the finished volume.

The challenge facing Vesalius was to ensure that his dissections did not distort the normal anatomy, and that his descriptions were accurate, precise and understandable. Dissection remains an effective teaching tool, but twentieth-century advances in imaging techniques mean that normal anatomy can now be comprehended in other forms, such as the 'slices' revealed by computed tomography (CT), and magnetic resonance imaging (MRI). Thus we now need the equivalent of the *Fabrica* with the anatomy presented, not by dissection, but in thin slices.

The Visible Human Project is the response to this need. The project was the brainchild of the US National Library of Medicine, but it was carried out at the University of Colorado Health Science Center at Denver. Two cadavers, one male and one female, were imaged using CT and MRI, at intervals of, respectively, 1 mm and 4 mm. They were then embedded in gelatin and sliced transversely, the male cadaver at 1 mm intervals, resulting in 1,878 cross-sections, and the female at 0.33 mm, producing 5,189 cross-sections. The slices were then photographed and digitized. This information was then stored on just two CDs, by compressing the data using the sophisticated MT-Wice system. The transverse sections are supplemented by images made in the coronal plane. Individual transverse cross-sections can be downloaded, and then annotated with PC- or Mac-compatible viewing software.

The Complete Visible Human relied on an awesome battery of technology, and on a multidisciplinary team of technical and anatomical experts, in stark contrast to the very personal achievements of Vesalius, his illustrators and his translators. However, the objective of both projects was to ensure that evidence, and not dogma, guided the anatomist and the clinician. To choose between them is invidious, but I have already

packed Richardson and Carman's splendid *Fabrica* to take to my desert island. □

Bernard Wood is in the Department of Anthropology, George Washington University, 2110 G Street, NW, Washington DC 20052, USA.

Undiplomatic immunity

The Generation of Diversity: Clonal Selection Theory and the Rise of Molecular Immunology

by Scott H. Podolsky and Alfred Tauber
Harvard University Press: 1998. 508pp. \$75, £46.95

Fred S. Rosen

The good deeds of the US Immigration and Naturalization Service often go unnoticed. Before he could be deported because of the expiry of his visa, Susumu Tonegawa moved from La Jolla, California to the Basel Institute of Immunology in Switzerland. Under the constraints of his changed circumstances, Tonegawa dropped his work on the SV40 virus and took up a study of the immunoglobulin genes. His landmark work in this field, which was to earn him a Nobel prize in 1987, settled the religious wars then raging in the immunology community.

After the Second World War the instructive theory of antibody formation, which starred the role of antigen in moulding the antibody response, was no longer tenable. In 1957, Sir Frank Macfarlane Burnet wrote his historic treatise on the clonal selection theory, a more biological view of antibody generation that superseded all instructive theories. Burnet's proposal gave rise to a Thomistic epistemology in which revealed truth and natural reason were meant to converge and explain the generation of diversity (GOD) in the humoral immune system. They didn't.

The problem of antibody diversity was so compelling that scientists from many fields were drawn to attempt an explanation. Leo Szilard, Sydney Brenner, Francis Crick and many others all got it wrong. Eventually, Philip Leder calculated that the humoral immune system could generate 18 million different specificities. Was all this information encoded in the germ line or could somatic recombination explain GOD? In the end Tonegawa showed definitively that the latter explanation was correct. The complete story was fleshed out by the important contributions of Leder, Bernard Mach, David Baltimore and others.

Scott Podolsky and Alfred Tauber have written a dense and exhaustive scholarly treatise about the history of GOD. They explore the intellectual setting in which the Tonegawa experiment was performed by reviewing original literature and interviewing the major actors in the drama. This is not

reading material for lay people, but everyone interested in the history of immunology and genetics should read this book.

The authors paraphrase the question of Charles A. Janeway, Jr: what is the purpose of the humoral immune system? He should know the answer: not much. Males with X-linked agammaglobulinaemia who can make no antibodies, and to whom GOD is inaccessible, need only a few dozen of Leder's 18 million specificities to keep them healthy. Fortunately, the principles that applied to immunoglobulins are also pertinent to the T-cell antigen receptor; that is what really matters. □

Fred S. Rosen is in the Department of Pediatrics, Harvard Medical School, Longwood Avenue, Boston MA 02115, USA.

Taking the mammoth steppes

Wildlife of the Tibetan Steppe

by George B. Schaller
University of Chicago Press: 1998. 384pp.
\$55, £43.95

Valerius Geist

Wildlife of the Tibetan Steppe is an account of George Schaller's investigations into the origins and ecology of the Tibetan megafauna, and of his dealings with the Chinese authorities in attempting to delineate wildlife reserves and formulate appropriate conservation policy. It begins with a most readable synthesis of the historical factors that shaped the Tibetan plateau, followed by chapters on large herbivores, carnivores and small mammals. He goes on to explore their feeding ecology, and analyse the phylogeny of Tibetan bovids using first morphological and molecular factors, and then behavioural components. There follows a very insightful chapter on the desperately poor nomadic people he encountered, and their attitude to wildlife. Finally, he makes recommendations on its conservation.

Why read about Tibetan wildlife? A specialist might be tantalized by the similarities between this fauna and that inhabiting the late Pleistocene "mammoth steppe", as the Alaskan palaeontologist Dale Guthrie aptly named it. In fact, this was a focus of the only comparable work to Schaller's, the dissertation on the zoogeography and ecology of the Tibetan plateau by Ernst Schäfer, unfortunately unpublished and in German (*Tiergeographisch-oekologische Studien über das Tibetische Hochland*, Munich, 1941). Before the Second World War, Schäfer was co-leader of the two US Brook-Dolan expeditions to eastern Tibet, and later of a German expedition to Lhasa. He was made an SS officer, which caused him grief during the war, and made him *persona non grata* widely there-



Well spotted: *Wildlife of the Tibetan Steppe* documents rare megafauna including the snow leopard.

after. Yet Schäfer was a brilliant scholar, and I regret that his synthesis reached Schaller's hands too late to be critically examined and integrated.

In the technical sections of the book, Schaller makes significant contributions to an understanding of the origins and ecology of Tibetan wildlife that will thrill specialists. I have difficulties with a few points, but these become mere quibbles between experts, for Schaller's book is much more than an ecological synthesis. It is a quest for conservation, a case history by a very brave and capable man, driven by no small passion to prevent the tragedy of extinction that looms over Tibet's fauna. His book touches not only the mind but also the heart, and in the context of conservation and the future it raises questions to torture the soul. In intertwining biology with conservation, Schaller's is a book of the times, one most revealing about the decay of our Earth's biota and the desperate clawing for answers to stop and reverse it. Schaller lays bare his approach and in so doing invites discussion about how better to conserve biodiversity in the context of non-Western cultures and overflowing human populations.

Schaller, ever mindful that he worked as a guest of China, suggests measures to conserve wildlife. Among these are suggestions that wildlife be used economically for its products, or via tourism, but with benefits for and the meaningful involvement of the local population. This reflects earlier discussions in conservation circles, but takes no account of controversy or other models. Here lies deep irony, because Schaller is an American biologist: his education was a product of policies that returned North American wildlife from the edge of extinction in the greatest environmental success story of the century.

Only 100 years ago, North America's wildlife was in a state of depletion surpassing that in Tibet today. The 'tragedy of the commons' had run its sorry course. Yet wildlife not only returned from near extinction, but

continues to rise in numbers. Overpopulation of deer, elk and geese is causing grief; visitors to large tracts of public land need to consider how to avoid attacks by bears. Hunting seasons and bag limits can be extremely generous. Moreover, Americans and Canadians reap a rich economic reward from their public wildlife, from which the private sector creates wealth and jobs.

How is it that these rich, powerful, heavily industrialized nations have made wildlife abundant? How is it that the North American system of wildlife conservation has not only sustained wildlife for more than eight decades, but increased its numbers, despite economic exploitation?

Schaller suggests scientific research as an integral part of the recovery of Tibetan wildlife. This happens to be a fundamental tenet of North American conservation, often called the 'Roosevelt doctrine'. Why then not speak of other tenets of North American wildlife conservation? These include the prohibition of markets in dead wildlife, the allocation of wildlife by law and not by the marketplace, and organized grassroots participation in the management of wildlife.

The genius of North American wildlife conservation was to draw in the common folk, making them enthusiastic and effective supporters of wildlife. They, in a capitalist society, made wildlife a great Public Good. How was that done, and are there no lessons in this? Of course, there are also conservation problems in North America. Many academics and natural resource managers have failed to learn from their own conservation successes, as proved by, among other cases, the tragic state of coastal fish stocks.

Wildlife of the Tibetan Steppe will long remain a unique, important source of biological, but also sociological, insights and challenges. I found it well written and difficult to put down. □

Valerius Geist is in the Faculty of Environmental Design, University of Calgary, and at PO Box 1294, Station A, Port Alberni, BC, Canada V9Y 7M2.

Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16^{INK4a}

Alicia A. Russo*, Lily Tong†, Jie-Oh Lee†, Philip D. Jeffrey* & Nikola P. Pavletich†

* Cellular Biochemistry and Biophysics Program and † Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA

The cyclin-dependent kinases 4 and 6 (Cdk4/6) that control the G1 phase of the cell cycle and their inhibitor, the p16^{INK4a} tumour suppressor, have a central role in cell proliferation and in tumorigenesis. The structures of Cdk6 bound to p16^{INK4a} and to the related p19^{INK4d} reveal that the INK4 inhibitors bind next to the ATP-binding site of the catalytic cleft, opposite where the activating cyclin subunit binds. They prevent cyclin binding indirectly by causing structural changes that propagate to the cyclin-binding site. The INK4 inhibitors also distort the kinase catalytic cleft and interfere with ATP binding, which explains how they can inhibit the preassembled Cdk4/6–cyclin D complexes as well. Tumour-derived mutations in INK4a and Cdk4 map to interface contacts, solidifying the role of CDK binding and inhibition in the tumour suppressor activity of p16^{INK4a}.

A common event in tumorigenesis is the deregulation of the eukaryotic cell cycle at the G1–S transition point, when the commitment to complete the cell cycle is made. So prevalent is this event that the vast majority of cancer cases have alterations in at least one of the main regulators of this transition^{1,2}. These include the retinoblastoma (Rb) tumour suppressor, which blocks the G1–S transition³; the cyclin-dependent kinases 4 and 6 (Cdk4/6), which drive G1 progression and phosphorylate and inactivate Rb⁴; the cyclin D oncoprotein, which activates Cdk4/6 (ref. 5); and the p16^{INK4a} tumour suppressor, which inhibits Cdk4/6 (ref. 6).

Alterations to p16^{INK4a} in cancer are particularly common^{2,7,8}, occurring in ~80% of cases with certain tumour types⁹. They can either be inherited, as in familial melanoma^{10,11}, or occur somatically in sporadic tumours⁹. Solidifying its role in tumorigenesis, deletion of p16^{INK4a} in mice results in the development of spontaneous tumours at a young age, and high sensitivity to carcinogenic treatments¹².

Central to the tumour-suppressor effects of p16^{INK4a} is its ability

to arrest cell growth at G1 by binding to and inhibiting Cdk4/6 (refs 6, 13). Tumour-derived p16^{INK4a} mutants, which lack the ability to arrest cell growth at G1 and revert cell transformation by activated oncogenes¹⁴, are also defective in Cdk4/6 binding and inhibition^{13,15–19}. Similarly, a Cdk4 mutant isolated from familial melanoma is refractory to p16^{INK4a} binding and inhibition and retains the ability to phosphorylate Rb^{20,21}.

p16^{INK4a} is the prototype of a family of Cdk inhibitors, specific for Cdk4/6, that also includes p15^{INK4b} (ref. 22), p18^{INK4c} (ref. 23) and p19^{INK4d} (refs 23, 24). Most observations suggest that these inhibitors act by blocking the association of Cdk4/6 with cyclin D, thus preventing the activation of the kinase. For example, in cells with high levels of endogenous p16^{INK4a}, Cdk4 is associated only with p16^{INK4a} and not with cyclin D (refs 23, 25–27); and in assembly reactions, p16^{INK4a} prevents the binding of Cdk4/6 to cyclin D (refs 24, 27, 28). *In vitro*, however, p16^{INK4a} can bind to and inhibit the preassembled Cdk4/6–cyclin D complex without dissociating cyclin D (refs 6, 13, 23), pointing to a multi-pronged mechanism

Table 1 Statistics from the crystallographic analysis

Data set	p16 ^{INK4a} –Cdk6					p19 ^{INK4d} –Cdk6	
	Native	Hg λ ₁	Hg λ ₂	Hg λ ₃	Au	Hg + Au	Native
Wavelength (Å)	0.9080	1.0093	1.0060	0.9959	0.9080	0.9080	1.0060
Beam line	CHESS A1	NSLS-X4a	NSLS-X4a	NSLS-X4a	CHESS A1	CHESS A1	NSLS-X4a
Resolution (Å)	3.4	3.9	3.9	3.9	4.0	3.8	2.8
Observations	68633	52515	51741	51960	22733	23460	88741
Unique reflections	12443	8629	8553	8553	7594	8394	33400
Data coverage (%)	97.7	97.9	98.1	97.9	96.8	89.5	96.9
R _{sym} (%)	6.8	7.1	7.4	7.6	6.6	8.7	5.5
MIR/MAD analysis:							
Phasing power	–	1.8	1.9	1.9	1.0	2.2	–
R _{cullis}	–	0.64	0.63	0.62	0.85	0.58	–
Anomalous R _{cullis}	–	0.89	0.86	0.88	–	–	–
Refinement statistics:							
	Resolution (Å)	Reflections (F > 1σ)	Protein atoms	R-factor (%)	R-free (%)	R.m.s. deviations	
						Bonds (Å)	Angles (°)
							B-factor (Å ²)
p16 ^{INK4a} –Cdk6	10.0–3.4	11109	3311	22.7	33.0	0.012	1.9
p19 ^{INK4d} –Cdk6	10.0–2.8	27794	3397*	25.5	30.8	0.017	2.1
							3.8

* Strict non-crystallographic symmetry was used throughout refinement; the number of atoms corresponds to only one of the two complexes in the asymmetric unit.
 $R_{\text{sym}} = \sum_i \sum_j |I_{ij} - \langle I \rangle| / \sum_i \sum_j I_{ij}$ for the intensity (I) of observations of reflection h . Phasing power = $\langle F_o \rangle / E$, where $\langle F_o \rangle$ is the r.m.s. heavy-atom structure factor and E is the residual lack of closure error. R_{cullis} is the mean residual lack of closure error divided by the dispersive or anomalous difference. Figure of merit = $|F(hkl)_{\text{obs}}| / |F(hkl)|$. R -factor = $\sum |F_o - F_c| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively. R -free = R -factor calculated using 5% of the reflection data chosen randomly and omitted from the start of refinement. R.m.s. deviations are deviations from ideal geometry and root mean square variation in the B -factor of bonded atoms.

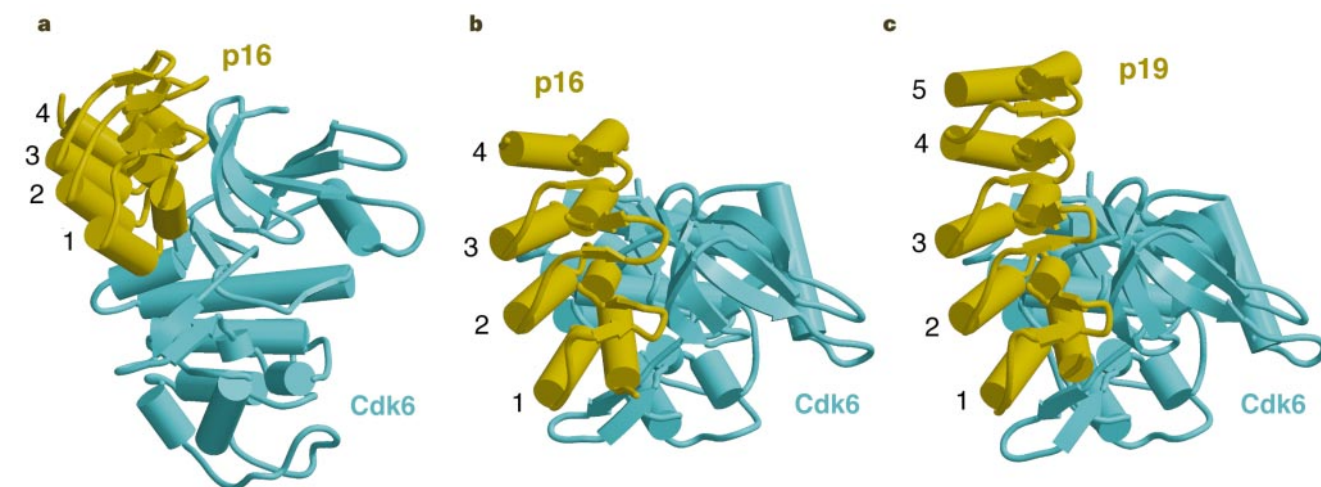


Figure 1 The INK4 inhibitors bind to one side of the catalytic cleft of Cdk6, and interact with both the N and C lobes. **a**, **b**, Schematic representation of the p16-Cdk6 complex in two orthogonal views, with the ankyrin repeats numbered. **c**,

View of the p19-Cdk6 complex in an orientation equivalent to that in **b**. Figures were made with the programs MOLSCRIPT⁵⁰ and RASTER3D⁵¹.

of action. These observations distinguish the INK4 family from the Kip/Cip family of CDK inhibitors, which have a broader CDK preference, bind the combined CDK-cyclin complex, and inhibit in part by blocking ATP binding²⁹.

The structures of the Cdk6-p16^{INK4a} and Cdk6-p19^{INK4d} complexes, reported here (Table 1), reveal major structural distortions in the cyclin-binding site as well as the ATP-binding site of Cdk6. These structural features help to explain how the INK4 inhibitors can block both the CDK activation process as well as the preactivated Cdk4/6-cyclin D complex.

Overall structure of the complex

The overall structure of p16-bound Cdk6 consists of an amino-terminal lobe rich in β -sheet (N lobe, residues 1–103), a larger α -helical carboxy-terminal lobe (C lobe, residues 104–301) and a catalytic cleft at the junction of the two lobes (Fig. 1a). In this general respect, p16-bound Cdk6 is similar to Cdk2 (50% sequence identity³⁰), and to other eukaryotic protein kinases³¹.

p16^{INK4a} (henceforth p16) and p19^{INK4d} (henceforth p19) consist of four and five ankyrin repeats, respectively. This ~30 amino acid structural motif, present in diverse proteins, has a structure that resembles the letter 'L' (ref. 32) with a pair of antiparallel helices forming the stem and a β -hairpin forming the base. As seen in the structures of monomeric p19 (ref. 33), p16 (ref. 34), and p18^{INK4c} (ref. 35), the L-like repeats stack to give an extended concave surface (Fig. 1a–c). Cdk6-bound p16 and p19 have very similar structures (1.0 Å r.m.s. deviation in C α positions), and they do not change significantly upon binding to Cdk6 as indicated by an r.m.s. deviation of 1.3 Å in C α positions between Cdk6-bound p16 and free p18 (ref. 35).

p16 binds to one side of the catalytic cleft, opposite where the cyclin would bind, and interacts with both the N and C lobes of Cdk6. The concave surface of p16 binds the N-lobe β -sheet, and the tips of the ankyrin repeats interact with the C lobe (Fig. 1a, b). Binding and recognition are mediated primarily by hydrogen-bond networks, with several of the residues that participate in these interactions being mutated in cancer (Figs 2, 3).

At the simplest level, p16-bound Cdk6 is inactive because it is not bound to a cyclin and is not phosphorylated. Previous structural studies of Cdk2 have shown that cyclin binding and phosphorylation are necessary to bring about activating conformational changes in two regulatory elements of the kinase^{30,36,37}: the PSTAIRE helix, pushed by the cyclin, has to move into the catalytic cleft to bring in

and realign active site residues^{30,36}; and the T loop has to undergo successive conformational changes, first upon cyclin binding^{30,36}, and then upon phosphorylation³⁷, to relieve the blockage of the catalytic cleft and to organize the polypeptide substrate-binding site.

Like inactive free Cdk2 (ref. 30), p16-bound Cdk6 does not have any of these activating changes; unlike free Cdk2, however, p16-bound Cdk6 has structural distortions that move it further away from the active structure. Pushed apart by p16, the N and C lobes of Cdk6 are twisted away from each other, distorting the cyclin-binding site and blocking the translocation of the PSTAIRE helix into the catalytic cleft (Fig. 4a, b). The ATP-binding site is distorted

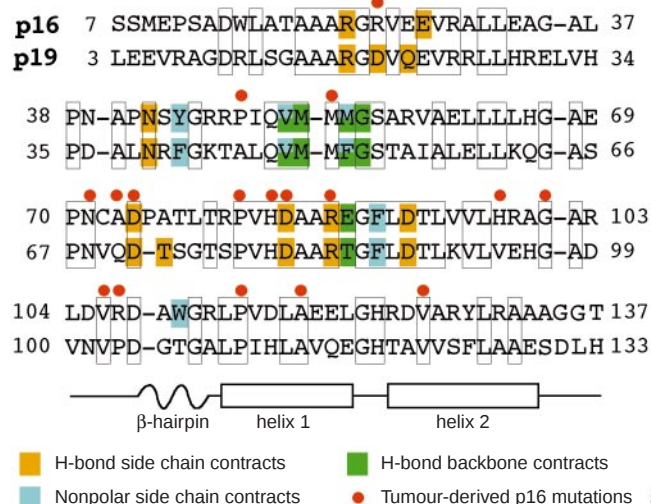


Figure 2 p16 and p19 use conserved regions to make overall conserved contacts with Cdk6. The four ankyrin repeats of p16 are aligned with the corresponding regions of p19 (the fifth p19 repeat is not shown). The ankyrin repeats have conserved secondary structure elements, except for the second ankyrin repeats of both p16 and p19, which have only one turn of helix 1. Although a much larger number of tumour-derived p16 mutants are known, those whose loss of function has not been determined are not shown. Among the differences in the p16-Cdk6 and p19-Cdk6 contacts, one involves Trp 110 of the fourth p16 repeat contacting a Cdk6 loop (Gly 36 nd Gly 37); in the p19 complex, the corresponding residue is a threonine that makes no contacts.

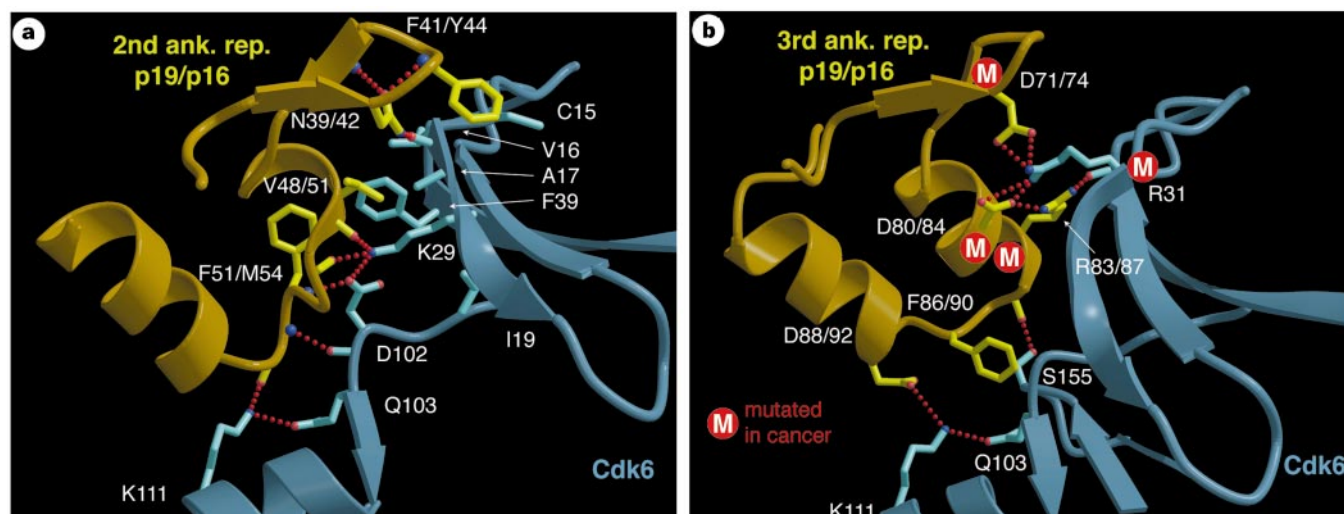


Figure 3 Conserved hydrogen-bond networks by the second and third ankyrin repeats of p16 and p19 have a central role in Cdk6 binding. **a**, Cross-section of the complex showing the contacts between Cdk6 and the second ankyrin repeat. **b**, Cross-section showing the third ankyrin repeat. Only the hydrogen-bond contacts conserved in the p16–Cdk6 and the p19–Cdk6 complexes are shown (red dotted

lines). The figures show the p19–Cdk6 complex, as the atomic positions of this 2.8-Å-refined model are of higher precision than those of the 3.4-Å-refined p16–Cdk6 complex. INK4 residues are labelled first with their p19 numbers and second with their p16 numbers, separated by a slash. The secondary structure labels of Cdk6 are according to Cdk2 (ref. 30).

because of contacts from p16, which binds adjacent to the catalytic cleft, and because of the misalignment of the N and C lobe. (Fig. 3a). This distortion would interfere with the productive binding of ATP, as also indicated by biochemical data in this study (Fig. 5b, c). The T loop, which seems to have significant structural freedom in CDKs, having been seen in three different places and conformations in the structures of free Cdk2 (ref. 30), cyclin A-bound Cdk2 (ref. 36), and cyclin A-bound phosphorylated Cdk2 (ref. 37), is farthest away from its expected active position, and is in the way of ATP binding as well (Fig. 6a, b).

The binding and mechanism of action of the INK4 family of CDK inhibitors is thus distinct from the Kip/Cip family of CDK inhibitors. However, it is remarkable that both families of inhibitors use the apparent structural flexibility of the kinase in their action. The Kip/Cip proteins remould the kinase N lobe, grossly changing the catalytic cleft shape²⁹, and the INK4 proteins both cause inhibitory structural changes and block activating structural changes.

Interactions at the interface

The p16–Cdk6 complex forms a continuous interface that buries a total of 2,200 Å² surface area. Roughly 60% of the buried area is contributed by interactions from the N lobe of Cdk6, with the remainder being from the C lobe. p16's most conserved second and third repeats contribute most of the interactions, whereas its first and fourth repeats make fewer interactions (Fig. 2). p19 binds similarly, except that its fourth repeat makes no significant contacts.

The binding to the N lobe is mediated by the concave surface of the four ankyrin repeats of p16, which contact the convex face of the twisted N-lobe β-sheet (β strands 1–3; Figs 1a, 3). The binding to the C lobe is mediated by the tips of the first three ankyrin repeats, which interact with the T loop (repeat 1) and with structural elements of the catalytic cleft (the α2 helix and β7–β8 strands; Figs 1a, 3).

The interface is mostly polar, relying on hydrogen-bond networks at solvent-excluded regions for binding and recognition. The interface is closely packed, as underscored by the participation of seven backbone groups, in addition to the fifteen side chain groups, in the hydrogen bond networks (Fig. 3). Most of the backbone hydrogen bonds involve the second repeat of p16, and in particular a loop (residues 52–54; Fig. 3a) that in the canonical ankyrin repeat

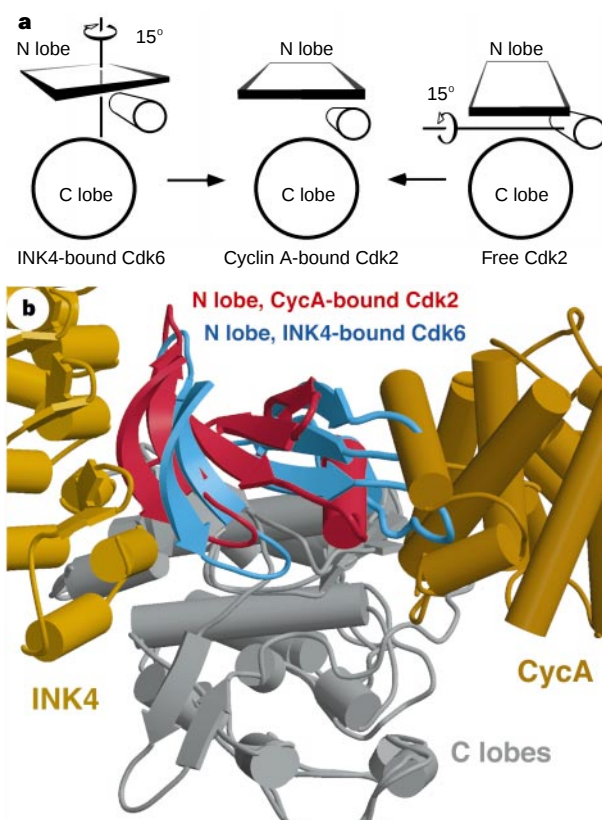


Figure 4 p16 locks the N and C lobes in a relative orientation not amenable to cyclin binding and activation. **a**, Schematic representation of the N and C lobes from p16-bound Cdk6, cyclin A-bound Cdk2 and free Cdk2. The C lobes are drawn as circles; the N-lobe β-sheets are indicated by rectangles and the PSTAIRE helices as cylinders. The axes through which rotations would bring the two lobes of p16-bound Cdk6 and of free Cdk2 into coincidence with those of the active cyclin A-bound Cdk2 are indicated. **b**, Composite figure showing the p16–Cdk6 and the cyclin A–Cdk2 complexes aligned by superimposing the C lobes of the two kinases.

structure would be an α -helix. This helix-to-loop change is a distinguishing structural feature of the INK4 family, as it is present in the structures of unbound p19 (ref. 33), p16 (ref. 34) and p18 (ref. 35) as well.

Contacts to the N and C lobes of Cdk6 are tightly coupled. For example, a hydrogen bond network of two side-chain and four backbone groups links the second ankyrin repeat of p16 to the N and C lobes of Cdk6 (centred on Lys 29 of Cdk6; Fig. 3a). Other hydrogen-bond networks link the third repeat to the N lobe (centred on Arg 31 of Cdk6; Fig. 3b), and the second and third repeats to the C lobe (centred on Lys 111 and Gln 103 of Cdk6; Fig. 3a, b). Mutation of the Cdk4 residues equivalent to Lys 29 or Arg 31 of Cdk6 abolishes p16 binding^{20,38}, underscoring the significance of these contacts.

Van der Waals contacts by hydrophobic residues are fewer (Figs 2, 3). One significant contact is made by the conserved Phe 90 of p16, which wedges between two loops that serve as the back wall of the catalytic cleft (residues 102–103 and 154–155 of Cdk6; Figs 3b, 5a). This contact is associated with the displacement of elements in the ATP-binding site.

Specificity for Cdk4/6. A subset of the Cdk6 residues that participate in the intermolecular hydrogen bond networks with p16 are conserved exclusively in Cdk4 and Cdk6 and are likely to be important determinants of the specificity of the INK4 family for these CDKs. Most of these residues are on the C lobe: Asp 102/Asp 97 in Cdk6 and Cdk4, respectively, are Ser/His/Ser in Cdk1, 2 and 3, respectively (Fig. 3a); Lys 111/Lys 106 are Ser/Ala/Ser (Fig. 3a, b); and Ser 155/Ser 150 are Asp/Thr/Glu (Fig. 3b). On the N lobe, the hydrogen-bonding groups (Lys 29 and Arg 31) are conserved in most CDKs, but there are several hydrophobic residues that make van der Waals contacts that show conservation correlated with INK4

specificity. The most notable residues are Ala 17/Ala 10 in Cdk6/Cdk4, which are Glu/Glu/Glu in Cdk1/2/3 respectively, and Phe 39/Phe 31, which are Val/Val/Leu (Fig. 3a).

It is likely that besides direct side chain contacts, the ability of Cdk6 to undergo conformational changes associated with p16 binding also contributes to the specificity of the INK4 proteins for this subfamily of CDKs. For example, a 10–25-residue C-terminal tail of CDKs, variable in sequence and length, is disordered in p16-bound Cdk6; in Cdk2 (ref. 30), the tail packs on a region that overlaps the p16 binding site of Cdk6.

Tumour-derived p16 and Cdk4 mutations at the interface. Solidifying the importance of CDK binding to the function of p16 as a tumour suppressor, all four of the amino acids that participate in open of the most extensive hydrogen-bond networks at the interface are mutated in cancer (Fig. 3b). This network contains Asp 74, Asp 84 and Arg 87 of p16, and Arg 31 of Cdk6, corresponding to Arg 24 of Cdk4, which is mutated in familial melanoma^{20,21}.

Other p16 residues that contact Cdk6 and are mutated in tumours include Glu 26 (ref. 34), which binds the T loop of Cdk6 (Fig. 2), and Asp 92 (ref. 34), which participates in another extended hydrogen-bond network with the C-lobe of Cdk6 (Fig. 3b), although these mutants have not yet been tested for loss of function.

The remaining tumour-derived missense mutations in p16 affect its structural integrity, as has been demonstrated by studies of its stability and aggregation state^{17,18,34}, and by the structures of free p19, p16 and p18 (refs 33–35).

p16 blocks cyclin binding indirectly

The binding of cyclin D to Cdk4/6 is likely to be similar to the binding of cyclin A to Cdk2 because the regions of cyclin A that contact Cdk2 (ref. 36) are conserved in cyclins. This is supported by

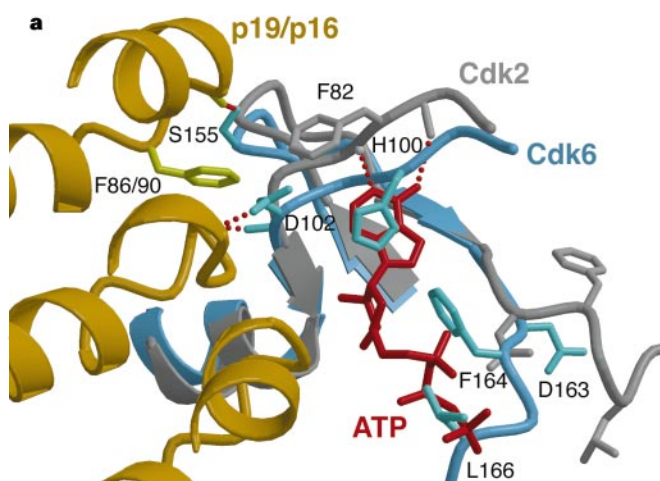
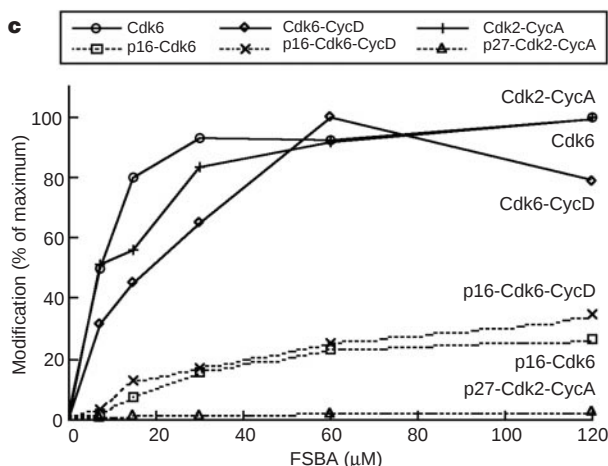
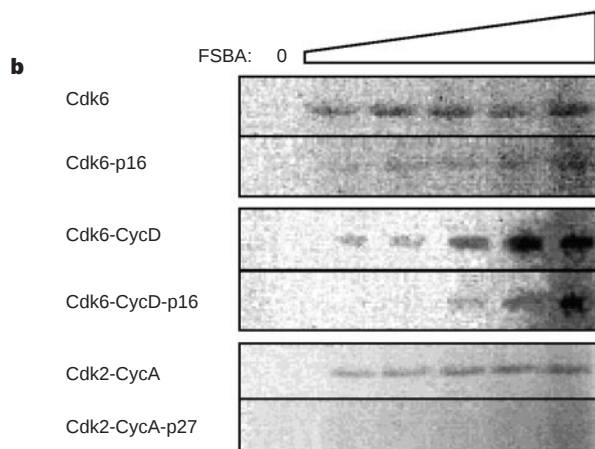


Figure 5 p16 alters the ATP-binding site of Cdk6 and interferes with ATP binding. **a**, Comparison of the ATP-binding sites from the p19-Cdk6 and the cyclin A-Cdk2-ATP complexes. The two complexes are superimposed as in Fig. 4b. The view is into the plane of the ATP-binding site in an orientation similar to that of Fig. 1b. For clarity, the N lobe, which would be above the plane of the figure, and most of the C lobe below the figure plane are not shown. **b**, [¹⁴C]FSBA concentration titration (0.0, 7.5, 15.0, 30.0, 60.0 or 120.0 μ M), showing that p16 reduces the reactivity of both Cdk6 and Cdk6-cyclin D for FSBA. The negative control, the Cdk2-cyclin-A complex bound to the p27^{Kip1/Cip2} cell cycle inhibitor that binds inside the catalytic cleft mimicking ATP²⁹, showed no FSBA reactivity under these conditions. The lysine residue (Lys 43 in Cdk6) to which FSBA would crosslink is not involved in any p16 interactions, being far from the p16-binding site and closer to the cyclin-binding site of the catalytic cleft. **c**, Percentage maximum modification was calculated by dividing [¹⁴C] incorporation for each reaction by the maximal incorporation for each kinase species in the absence of p16. The reactions containing p16 reached near-maximum modification levels typically at 500 μ M FSBA (not shown).



the ability of cyclin D to bind Cdk2 as well. In the Cdk2–cyclin A complex³⁶, the cyclin-binding site on the kinase consists of the β -sheet from the N lobe (15% of the CDK–cyclin interface area), the PSTAIRE helix (30%), the T loop (25%) and part of the C lobe (30%). In the p16-bound Cdk6, these structural elements differ from those of free Cdk2 or cyclin A-bound Cdk2 in their relative position or orientation. The presence of p16 would probably block some of the structural changes required for cyclin binding, and might present an activation barrier to others.

Misalignment of the N and C lobes. The relative orientations of the N and C lobes differ between p16-bound Cdk6 and cyclin A-bound Cdk2 by a 15° rotation about an axis perpendicular to the plane of the catalytic cleft (a ‘vertical’ axis in Fig. 4a). This axis of rotation passes through Val 76, roughly equidistant from the p16 and cyclin-binding sites, at the back of the cleft. The rotation, which is present in an identical form in the p19–Cdk6 complex, is tightly coupled to, and is probably a consequence of, the contacts that these INK4 proteins make to the two lobes (Fig. 3).

As a result of this rotation, portions of the N-lobe β -sheet at the putative cyclin-binding site undergo lateral displacements of 5–7 Å relative to the C lobe (Fig. 4b). In the presence of p16, restoration of cyclin-binding site would require structural changes to the N-lobe β -sheet. We think this is unlikely because the core of the N-lobe β -sheet is generally conserved in sequence and structure. For example, 59 of the 83 residues of the Cdk6 and Cdk2 N lobes can be superimposed with a C α r.m.s. deviation of 1.8 Å.

This rotation about a vertical axis in the p16-bound Cdk6 lobes is unlike that which free Cdk2 undergoes upon binding to cyclin A, where the N and C lobes rotate by 15° about a horizontal axis in the plane of the catalytic cleft (Fig. 4a). This rotation of the lobes in the Cdk2–cyclin A complex results in the opening of the catalytic cleft, making room for the translocation of the PSTAIRE helix into the cleft^{30,36}.

PSTAIRE helix translocation blocked. In both the p16–Cdk6 and p19–Cdk6 structures the PSTAIRE helix has high temperature factors and is not well ordered, although its overall position is well determined in the maps of the p16–Cdk6 complex derived from a combination of multiple isomorphous replacement (MIR) and multi-wavelength anomalous diffraction (MAD) methods. To reconstitute the cyclin-binding site, the PSTAIRE helix will have to move into the catalytic cleft relative to the C lobe. This translocation is indirectly antagonized by p16, because the misalignment of the N and C lobes places side chains in the way (Fig. 4b).

T loop is away from the cyclin-binding site. In both the p16 and p19-bound Cdk6 structures, the T loop (residues 162–182) exits the

catalytic cleft from the side where the INK4 inhibitor is bound, this being aided by its interactions with the uniquely aligned N and C lobes (Fig. 6a). This is in contrast to the Cdk2 structures, where the T loop exits the catalytic cleft closer to the cyclin-binding side (Fig. 6a). Immediately after it exits the catalytic cleft, the T loop forms a 10-residue β -hairpin, which is well ordered and rigid in the p19–Cdk6 structure (Fig. 6b), but is partly flexible in the p16–Cdk6 structure. The β -hairpin packs on the C lobe via hydrogen-bond and hydrophobic interactions, and also makes a set of hydrogen-bond contacts to p19 (Arg 168 of Cdk6 with Asp 20 and Gln 22 of p19; Fig. 2). It is remarkable that this is the fourth distinct T-loop conformation observed in CDK structures, and represents the largest displacement away from the T-loop structure observed in the fully active cyclin A-phosphorylated Cdk2 complex (with a maximal displacement of 35 Å compared to 21 Å for free Cdk2; ref. 36).

In the structure of mouse p19 bound to human Cdk6 (ref. 39), the T loop is observed in yet another conformation. It is not known which T-loop conformation is favoured in solution, as crystallization conditions, crystal packing forces and other effects can shift the equilibrium in the crystallization experiments. Nevertheless, the observed variety in T-loop conformations supports a role for the inherent flexibility of the T loop, as seen in the Cdk2 structures as well, in keeping the CDK inactive.

Distortion of the ATP-binding site

In both the p16- and p19-bound Cdk6 structures, an internal catalytic cleft loop (residues 99–102) that would otherwise make backbone hydrogen bonds to the N6 and N1 groups of the ATP purine³⁶ is dislodged by an average of 2.5 Å by direct contacts from p16 (Fig. 5a). N- and C-lobe side chains that would contact the ATP purine and phosphate groups as seen in the Cdk2 structures^{30,36} are partly out of register owing to the misalignment of the two lobes. In addition, side chains from the dislodged internal catalytic cleft loop (His 100) and from the T loop (Phe 164 and Leu 166) as it exits from the cleft have invaded portions of the space where the ATP purine, sugar and phosphate groups would be (Fig. 5a). These features of the catalytic cleft would not entirely preclude ATP from binding the way the Kip/Cip family of inhibitors do, as there is room remaining, especially if some side chains twist out of the way. But they would reduce the affinity of the kinase for ATP, and could well misorient any bound ATP with respect to catalysis. This is supported by binding studies of the ATP analogue FSBA (*p*-fluorosulphonylbenzoyl 5'-adenosine), which crosslinks covalently to an active-site lysine residue conserved in protein kinases^{40,41}. Figure 5b and c

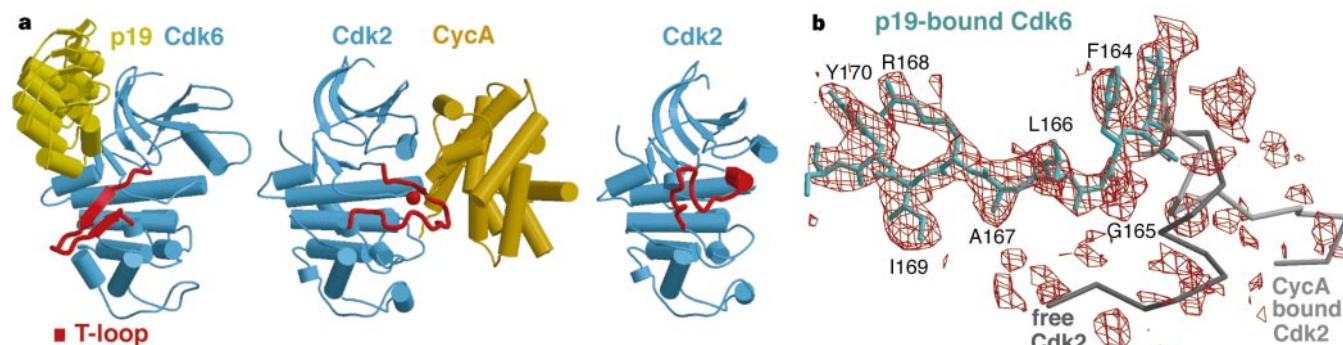


Figure 6 The T loop has large displacements from the active conformation compared to Cdk2 structures. **a**, Comparison of the three structures oriented by their C lobes highlights the different conformations and positions of their T loops. Equivalent portions of the T loops between the conserved DFG and APE sequences are shown in red. The red sphere in cyclin-A-bound Cdk2 indicates the phosphate group on the Thr 160 of that kinase. **b**, Omit difference electron density of the p19–Cdk6 complex showing the portion of the T loop as it exits from

the catalytic cleft and the first strand of the β -hairpin (the second strand is below the plane of the figure). Model bias was removed by omitting the T loop (residues 162–184) and by subjecting the model to simulated annealing refinement from 4,000 K with ncs-strict in X-PLOR. The map was averaged iteratively with the program RAVE⁴⁸, and the resulting phases were used to construct the $F_o - F_c$ omit map. Only difference density in the vicinity of the T-loop atoms (including side-chain atoms from the two Cdk2 T loops) is shown. The map is contoured at 2.0 σ .

shows that p16 binding reduces the reactivity of Cdk6 for FSBA by a factor of ~20.

Inhibition of the Cdk4/6–cyclin-D complex by p16

It is not yet clear whether the structure of the kinase in the ternary p16–Cdk4/6–cyclin D complex^{6,13,23} would be closer to that of p16-bound Cdk6 or cyclin-A-bound Cdk2. Nevertheless, the p16–Cdk6–cyclin D complex has FSBA reactivity that is ~20-fold lower than that of the Cdk6–cyclin D complex (Fig. 5b, c). This indicates that when p16 is bound to the Cdk6–cyclin D complex, at least some of the distortions in the ATP-binding site are present and contribute to the inhibition of the Cdk6–cyclin D complex by p16.

One model for the binding of p16 to Cdk6–cyclin D is that p16 might still make most of the N-lobe contacts, as it has some detectable affinity for the isolated N lobe³⁴. This might be sufficient to bring about some of the changes in the ATP-binding site. This model is consistent with the report of a 20-residue p16 peptide with residual Cdk4 binding and inhibitory activity⁴². This peptide encompasses the residues that form the hydrogen-bond network with Arg 31 of the N lobe of Cdk6, which could serve as an initial binding point (Fig. 3b), and it also contains the Phe 90 residue, which has a central role in the dislocation of the ATP-interacting loop in the catalytic cleft (Figs 3b, 5a).

Implications for other Cdk4/6 complexes

The structural distortions seen in p16-bound Cdk6 raise the possibility that other macromolecular complexes of Cdk6 might also be antagonized by INK4 binding. We tested Cdc37, which might be required for Cdk–cyclin association in yeast⁴³ and might stabilize Cdk4 in mammalian cells⁴⁴. Binding studies *in vitro* with purified proteins (not shown) indicate that p16 both can block Cdc37 binding to Cdk6 and also can dissociate prebound Cdc37–Cdk6 complexes.

INK4 proteins also inhibit the binding of the Kip/Cip family of inhibitors to Cdk4/6 (ref. 45). The structure reveals that the two families of inhibitors have partly overlapping binding sites on the CDK (ref. 29).

Structural flexibility in CDK regulation

CDKs are among the most extensively and diversely regulated eukaryotic protein kinases, as they have to respond to a multitude of growth regulatory pathways that converge on the cell cycle. Structural studies of Cdk2 activation by cyclin A (ref. 36) and by phosphorylation³⁷, and of inhibition by p27^{Kip1/Cip2} (ref. 29) have revealed that all of these processes are associated with conformational changes. The structures of the p16–Cdk6 and p19–Cdk6 complexes strongly reinforce the concept that cyclin-dependent kinases possess an inherent structural flexibility that is central to their regulation. □

Methods

Protein overexpression and purification. Human Cdk6 (ref. 46) was overexpressed using a baculovirus vector in Hi5 (Invitrogen) insect cells grown in suspension in serum-free media (Sf900, Gibco). It was purified by affinity chromatography using a column of human p16, prepared as a glutathione *S*-transferase (GST)–p16 fusion in *Escherichia coli*. The affinity-purified GST–p16–Cdk6 complex was treated with thrombin at 4 °C to cleave the GST fusion protein, and was further purified by anion-exchange and gel-filtration chromatography. The complex was concentrated by ultrafiltration to typically 50 mg ml⁻¹ in a buffer of 25 mM Tris-HCl, 200 mM NaCl, 5 mM dithiothreitol (DTT), pH 7.5. The complex of human p19 bound to Cdk6 was prepared similarly, and was concentrated to 35 mg ml⁻¹. The overexpressed Cdk6 had a mass consistent with *N*-acetylation and lack of phosphorylation.

For the biochemical studies, Cdk6 was overexpressed as a GST–Cdk6 fusion protein in insect cells. After cleavage from the GST, Cdk6 was purified by anion-exchange and gel-filtration chromatographies in the same buffer as the p16–Cdk6 complex. The Cdk6–cyclin D complex was prepared by co-infecting

insect cells with separate baculoviruses expressing GST–Cdk6 and cyclin D1, then isolating, purifying and cleaving the GST–Cdk6–cyclin D complex as with free Cdk6. After cleavage from GST, free Cdk6, which was in large excess, was separated from the Cdk6–cyclin D complex by anion-exchange and gel-filtration chromatographies.

Crystallization and data collection. Crystals of the p16–Cdk6 complex were grown at 4 °C by the hanging-drop vapour-diffusion method, from 11% isopropanol, 40 mM Hepes-Na, 5 mM DTT, pH 7.5. They form in space group *P*₄₁₂₂ with *a* = *b* = 123.4 Å, *c* = 114.6 Å and contain one complex in the asymmetric unit. Heavy-atom soaks were performed in crystallization buffer lacking DTT with 0.4 mM thimerosal for 4 h, or 1 mM K₂Au(CN)₄ for 2 h, or 0.1 mM thimerosal and 2 mM K₂Au(CN)₄ for 4 h. Diffraction data were collected with crystals flash-frozen in crystallization buffer supplemented with 25% PEG400 at –170 °C.

Crystals of the p19–Cdk6 complex were grown from 26% saturated (NH₄)₂(SO₄), 8–12% dioxane, 100 mM MES-Na, 5 mM DTT, pH 6.5. They form in space group *C*2 with *a* = 98.4, *b* = 116.8, *c* = 132.2 Å, β = 110.4° and have two complexes in the asymmetric unit. The crystals were pretreated with glutaraldehyde (diffused into the drop from a reservoir of 30% glutaraldehyde) to reduce their tendency to crack and lose diffraction along *b** and *c**. They were then harvested in 30% saturated (NH₄)₂(SO₄), 10% dioxane, 100 mM MES-Na, 5 mM DTT, pH 6.5, supplemented with 20% glycerol just before flash freezing at –170 °C. Diffraction was strongly anisotropic, extending to 2.1 Å along *a**, compared to only ~3 Å along *c**.

Structure determination and refinement. The structure of the p16–Cdk6 complex was determined by a combination of MIR and MAD. Initial phases (mean figure of merit of 0.72) were calculated at 3.8 Å resolution with the program MLPHARE⁴⁷ and were improved by density modification with the program DM⁴⁷. A model of Cdk6 derived from the structure of Cdk2, and a model of p16 derived from the structure of p18 were readily built into the MIR/MAD maps with the program O (ref. 48). The model was refined with the programs X-PLOR⁴⁹ and REFMAC⁴⁷ to 3.4 Å. The refinement included the MIR/MAD phases as additional restraints, solvent correction, and grouped B factors (two per residue). The final model contains residues 10–134 of p16, and 10–48 and 57–301 of Cdk6. Residues 48–72 and 167–181 have high temperature factors; we presume they are partly disordered.

The structure of the p19–Cdk6 complex was determined by molecular replacement using the p16–Cdk6 structure as the search model. Anisotropic *B* factor refinement with X-PLOR resulted in a large correction to the data (*B*₁₁ = –29.0, *B*₂₂ = –8.0, *B*₃₃ = 37.0, *B*₁₃ = 17.9 Å²), which significantly improved the free *R*. The initial 2*F*_o – *F*_c maps were averaged iteratively with the program RAVE⁴⁸ to a correlation coefficient of 0.78. Throughout refinement the two complexes in the asymmetric unit were kept identical by applying strict non-crystallographic symmetry (ncs) in X-PLOR, as there was no evidence of significant positional differences between the two complexes when the strict-ncs constraints were removed. After restrained temperature factor refinement, simulated annealing omit maps (X-PLOR) averaged by RAVE were used to confirm every part of the complex systematically. Solvent correction was applied with X-PLOR. The final model contains residues 6–160 of p19, and residues 11–48, 57–87 and 92–301 of Cdk6, and has no water molecules.

Labelling with [¹⁴C]FSBA. Reactions, performed in accordance with published procedures⁴¹, contained 10.0 μM protein, 50 mM NaCl, 25 mM Hepes-Na, pH 7.5, [¹⁴C]FSBA (5′-*p*-[adenine-8-¹⁴C]fluorosulphonylbenzoyl adenosine; NEN–Du Pont), and 10% DMSO (from the added FSBA), were incubated at 30.0 °C for 15 min, then resolved by SDS–PAGE. The gels were dried without fixing, exposed to a phosphorimager plate, and ¹⁴C incorporation was quantitated by integration of the bands with background correction. For each kinase species, the reactions with and without p16 were done simultaneously and were repeated at least twice.

Received 15 May; accepted 10 July 1998.

- Sherr, C. J. Cancer cell cycles. *Science* **274**, 1672–1677 (1996).
- Hall, M. & Peters, G. Genetic alterations of cyclin, cyclin-dependent kinases, and cdk inhibitors in human cancer. *Adv. Cancer Res.* **68**, 68–108 (1996).
- Weinber, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
- Sherr, C. J. G1 phase progression cycling on cue. *Cell* **79**, 551–555 (1994).
- Hunter, T. & Pines, J. Cyclins and cancer II: cyclin D and cdk inhibitors come of age. *Cell* **79**, 573–582 (1994).

6. Serrano, M., Hannon, G. J. & Beach, D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* **366**, 704–707 (1993).
7. Kamb, A. *et al.* A cell cycle regulator potentially involved in genesis of many tumor types. *Science* **264**, 436–440 (1994).
8. Nobori, T. *et al.* Deletions of the cyclin-dependent kinase-4 inhibitor gene in multiple human cancers. *Nature* **368**, 753–756 (1994).
9. Caldas, C. *et al.* Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nature Genet.* **8**, 27–32 (1994).
10. Hussussian, C. J. *et al.* Germline p16 mutations in familial melanoma. *Nature Genet.* **8**, 15–21 (1994).
11. Kamb, A. *et al.* Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. *Nature Genet.* **8**, 23–26 (1994).
12. Serrano, M. *et al.* Role of the INK4a locus in tumor suppression and cell mortality. *Cell* **85**, 27–37 (1996).
13. Koh, J., Enders, G. H., Dynlacht, B. D. & Harlow, E. Tumour-derived p16 alleles encoding proteins defective in cell-cycle inhibition. *Nature* **375**, 506–510 (1995).
14. Serrano, M., Gomez-Lahoz, E., DePinho, R. A., Beach, D. & Bar-Sagi, D. Inhibition of ras-induced proliferation and cellular transformation by p16INK4. *Science* **267**, 249–252 (1995).
15. Yang, R., Gombart, A. F., Serrano, M. & Koeffler, H. P. Mutational effects on the p16INK4a tumor suppressor protein. *Cancer Res.* **55**, 2503–2506 (1995).
16. Ranade, K. *et al.* Mutations associated with familial melanoma impair p16INK4 function. *Nature Genet.* **10**, 114–116 (1995).
17. Parry, D. & Peters, G. Temperature-sensitive mutants of p16^{CDKN2} associated with familial melanoma. *Mol. Cell. Biol.* **16**, 3844–3852 (1996).
18. Zhang, B. & Peng, Z. Defective folding of mutant p16(INK4) proteins encoded by tumor-derived alleles. *J. Biol. Chem.* **271**, 28734–28737 (1996).
19. Arap, W., Knudsen, E. S., Wang, J. Y., Cavenee, W. K. & Huang, H. J. Point mutations can inactivate in vitro and in vivo activities of p16(INK4a)/CDKN2A in human glioma. *Oncogene* **14**, 603–609 (1997).
20. Wolfel, T. *et al.* A p16^{INK4a}-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science* **269**, 1281–1284 (1995).
21. Zuo, L. *et al.* Germline mutations in the p16^{INK4a} binding domain of CDK4 in familial melanoma. *Nature Genet.* **12**, 97–99 (1996).
22. Hannon, G. J. & Beach, D. p15^{INK4B} is a potential effector of TGF β -induced cell cycle arrest. *Nature* **371**, 257–261 (1994).
23. Hirai, H., Roussel, M. F., Kato, J. Y., Ashmun, R. A. & Sherr, C. J. Novel INK4 proteins, p19 and p18, are specific inhibitors of the cyclin D-dependent kinases CDK4 and CDK6. *Mol. Cell. Biol.* **15**, 2672–2681 (1995).
24. Guan, K. L. *et al.* Isolation and characterization of p19^{INK4d}, a p16-related inhibitor specific to CDK6 and CDK4. *Mol. Biol. Cell* **7**, 57–70 (1996).
25. Xiong, Y., Zhang, H. & Beach, D. Subunit rearrangement of the cyclin-dependent kinases is associated with cellular transformation. *Genes Dev.* **7**, 1572–1583 (1993).
26. Li, Y., Nichols, M. A., Shay, J. W. & Xiong, Y. Transcriptional repression of the D-type cyclin-dependent kinase inhibitor p16 by the retinoblastoma susceptibility gene product pRb. *Cancer Res.* **54**, 6078–6082 (1994).
27. Parry, D., Bates, S., Mann, D. J. & Peters, G. Lack of cyclin D–Cdk complexes in Rb-negative cells correlates with high levels of p16INK4/MTS1 tumour suppressor gene product. *EMBO J.* **14**, 503–511 (1995).
28. Lukas, J. *et al.* Retinoblastoma-protein-dependent cell-cycle inhibition by the tumour suppressor p16. *Nature* **375**, 503–506 (1995).
29. Russo, A. A., Jeffrey, P. D., Patten, A. K., Massagué, J. & Pavletich, N. P. Crystal structure of the p27^{Kip1} cyclin-dependent-kinase inhibitor bound to the cyclin A–Cdk2 complex. *Nature* **382**, 325–331 (1996).
30. De Bondt, H. L. *et al.* Crystal structure of cyclin-dependent kinase 2. *Nature* **363**, 595–602 (1993).
31. Knighton, D. R. *et al.* Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. *Science* **253**, 407–413 (1991).
32. Gorina, S. & Pavletich, N. P. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science* **274**, 1001–1005 (1996).
33. Luh, F. Y. *et al.* Structure of the cyclin-dependent kinase inhibitor p19^{INK4d}. *Nature* **389**, 999–1003 (1997).
34. Byeon, I. *et al.* Tumor suppressor p16^{INK4A}: determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4. *Mol. Cell* **1**, 421–431 (1998).
35. Venkataramani, R., Swaminathan, K. & Marmostein, R. Crystal structure of the CDK4/6 inhibitory protein p18^{INK4c} provides insights into ankyrin-like repeat structure/function and tumor-derived p16^{INK4} mutations. *Nature Struct. Biol.* **5**, 74–81 (1998).
36. Jeffrey, P. D. *et al.* Mechanism of CDK activation revealed by the structure of cyclin A–CDK2 complex. *Nature* **376**, 313–320 (1995).
37. Russo, A., Jeffrey, P. D. & Pavletich, N. P. Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nature Struct. Biol.* **3**, 696–700 (1996).
38. Coleman, K. G. *et al.* Identification of CDK4 sequences involved in cyclin D1 and p16 binding. *J. Biol. Chem.* **272**, 18869–18874 (1997).
39. Brotherton, D. H. *et al.* Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell cycle inhibitor p19^{INK4d}. *Nature* **395**, 244–250 (1998).
40. Zoller, M. J., Nelson, N. C. & Taylor, S. S. Affinity labeling of cAMP-dependent protein kinase with p-Fluorosulfonylbenzoyl adenosine. *J. Biol. Chem.* **256**, 10837–10842 (1981).
41. Atherton-Fessler, S., Parker, L. L., Geahlen, R. L. & Piwnicka-Worms, H. Mechanisms of p34^{cdc2} regulation. *Mol. Cell. Biol.* **13**, 1675–1685 (1993).
42. Fahraeus, R., Paramio, J. M., Ball, K. L., Lain, S. & Lane, D. P. Inhibition of pRb phosphorylation and cell-cycle progression by a 20-residue peptide derived from p16^{CDKN2/INK4A}. *Curr. Biol.* **6**, 84–91 (1996).
43. Gerber, M. R., Farrell, A., Deshaies, R. J., Herskowitz, I. & Morgan, D. O. Cdc37 is required for association of the protein kinase Cdc28 with G1 and mitotic cyclins. *Proc. Natl Acad. Sci. USA* **92**, 4651–4655 (1995).
44. Stepanova, L., Leng, X., Parker, S. B. & Harper, J. W. Mammalian p50Cdc37 is a protein kinase-targeting subunit of Hsp90 that binds and stabilizes Cdk4. *Genes Dev.* **10**, 1491–1502 (1996).
45. Reynisdottir, I., Polyak, K., Iavarone, A. & Massagué, J. Kip/Cip and Ink4 Cdk inhibitors cooperate to induce cell cycle arrest in response to TGF- β . *Genes Dev.* **9**, 1831–1845 (1995).
46. Meyerson, M. & Harlow, E. Identification of G1 kinase activity for cdk6, a novel cyclin D partner. *Mol. Cell. Biol.* **14**, 2077–2086 (1994).
47. Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
48. Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
49. Brunger, A. T. *X-PLOR, a System for Crystallography and NMR* (Yale Univ. Press, New Haven, CT, 1991).
50. Kraulis, P. J. Molscript: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
51. Merrit, E. A. & Murphy, M. E. Raster3D Version 2.0: a program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).

Acknowledgements. We thank S. Geromanos and H. Erdjument-Bromage of the Sloan-Kettering Microchemistry Facility for N-terminal sequence and mass spectroscopic analyses; H. Chou and E. D. Harlow for baculovirus vectors expressing Cdk6, GST–Cdk6 and cyclin D1; R. Marmorestein for the crystallographic coordinates of p18; C. Ogata of the National Synchrotron Light Source X4A beam line and the staff of the Cornell High Energy Synchrotron Source MacChess for help with data collection; and C. Murray for administrative help. Supported by the Howard Hughes Medical Institute, the NIH, the Pew Charitable Trusts, the Dewitt Wallace Foundation, and the Samuel and May Rudin Foundation.

Correspondence and requests for materials should be addressed to N.P.P. (e-mail: Nikola@xray2.mskcc.org). Coordinates have been deposited with the Brookhaven Protein Data Bank (accession codes 1BI7 and 1BI8.).

Crystal structure of the complex of the cyclin D-dependent kinase Cdk6 bound to the cell-cycle inhibitor p19^{INK4d}

Deborah H. Brotherton*, Venugopal Dhanaraj*, Scott Wick†, Leonardo Brizuela†, Peter J. Domaille‡, Elena Volynik*, Xu Xu†, Emilio Parisini*, Brian O. Smith*, Sharon J. Archer‡, Manuel Serrano§, Stephen L. Brenner‡, Tom L. Blundell* & Ernest D. Laue*

* Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK

† Mitotix Inc., One Kendall Square, Building 600, Cambridge, Massachusetts 02139, USA

‡ Du Pont Pharmaceutical Company, Box 80353, Wilmington, Delaware 19880-0353, USA

§ Department of Immunology and Oncology, Centro Nacional de Biotecnología, Campus de Cantoblanco, Madrid E-28049, Spain

The crystal structure of the cyclin D-dependent kinase Cdk6 bound to the p19^{INK4d} protein has been determined at 1.9 Å resolution. The results provide the first structural information for a cyclin D-dependent protein kinase and show how the INK4 family of CDK inhibitors bind. The structure indicates that the conformational changes induced by p19^{INK4d} inhibit both productive binding of ATP and the cyclin-induced rearrangement of the kinase from an inactive to an active conformation. The structure also shows how binding of an INK4 inhibitor would prevent binding of p27^{Kip1}, resulting in its redistribution to other CDKs. Identification of the critical residues involved in the interaction explains how mutations in Cdk4 and p16^{INK4a} result in loss of kinase inhibition and cancer.

Cyclin-dependent kinases (CDKs) are important in controlling the eukaryotic cell cycle. CDKs are switched on and off at different times; Cdc2/cyclin B regulates entry and exit from mitosis, whereas the cyclin D-dependent kinases (Cdk4/cyclin D and Cdk6/cyclin D) and those of Cdk2 (Cdk2/cyclin E and Cdk2/cyclin A) regulate G1 progression and entry into the S phase of the cell cycle¹.

CDKs are partially activated by their regulatory cyclin subunits, which are periodically synthesized and degraded during the cell cycle. The cyclin/CDK complex is fully activated by phosphorylation of a threonine residue in the T-loop by CDK-activating kinase (CAK)². In addition, phosphorylation of tyrosine and/or threonine residue(s) in the phosphate-binding loop by Wee1 and Myt1 inhibits CDKs³. This inhibition can be reversed by the dual-specificity Cdc25 phosphatases⁴.

CDKs are further regulated by CDK inhibitors (CKIs), which induce cell-cycle arrest in response to different signals^{5,6}. In

proliferating cells, their redistribution between the different complexes coordinates the timing of activation of the different CDKs^{7,8}. Two classes of CKI have been identified. First, the Cip/Kip family members, which include p21^{Cip1}, p27^{Kip1} and p57^{Kip2}, inhibit all G1- and S-phase CDKs and are important in p53- and TGFβ-mediated cell-cycle arrest⁵. Second, the INK4 family, which is specific for Cdk4 and Cdk6 (ref. 9), has four known family members, p16^{INK4a}, p15^{INK4b}, p18^{INK4c} and p19^{INK4d} (refs 9–13). These bind in the absence or presence of cyclin D, but show different patterns of expression⁶. Their interactions with Cdk4 and Cdk6 regulate phosphorylation of the retinoblastoma protein (pRB), which leads to disruption of pRB/E2F complexes and activation of S-phase genes¹⁴. Changes in expression or mutations to cyclin D1, to Cdk4, to p16^{INK4a} and to pRB are all strongly implicated in cancer¹.

To help understand the regulation of cyclin D-dependent kinases, we have determined the three-dimensional structure of the p19^{INK4d}/Cdk6 complex by X-ray diffraction analysis (Fig. 1 and Table 1).

Overall structure of the complex

Cdk6, in common with other kinases, consists of two domains with the catalytic cleft lying between the two. The overall structure of Cdk6 is very similar to that of Cdk2 (refs 15–17). The amino-terminal domain (residues 5–100) comprises a five-stranded β-sheet that packs against the PLSTIRE helix α1 (PSTAIR in Cdk2). The cyclin interacts with helix α1, whose conformation switches to activate the kinase^{16,17}. The larger carboxy-terminal domain (residues 101–309) is mainly α-helical, with a small β-sheet (Fig. 2).

p19^{INK4d} comprises five ankyrin repeats¹⁸ (I–V), each consisting of an extended strand followed by a helix–loop–helix (HLH) motif and another extended strand. Consecutive repeats are linked by a series of β-turns between the N and C termini. The antiparallel helices in each repeat stack together, giving the protein an elongated L-shaped structure (Fig. 2). The four helical bundles form the long

Table 1 Data collection and statistics from the crystallographic analysis

Data set	Native	Native
Resolution (Å)	15.0–2.1	18.54–1.90
Observations	326,706	166,218
Unique reflections	31,989	41,965
Data coverage (%)	97.8	98.5
R _{sym} (%)	9.6	5.4
Refinement		
Resolution range (Å)		18.54–1.90
R-factor/Free R-factor (all data)		20.0/25.3
Number of atoms		3,830
Number of waters		286
R.m.s.d. bond lengths (Å)*		0.024
R.m.s.d. bond angles (deg)*		2.3
R.m.s.d. B-factors* backbone/side chain (Å ²)		1.8/3.3

R_{sym} = $\sum_i \sum_j |I_{ij} - \bar{I}| / \sum_i \sum_j I_{ij}$ for the intensity (I) of i observations of reflection h.

R-factor = $\sum |F_o - F_c| / \sum |F_o|$, where F_o and F_c are the observed and calculated structure factors, respectively.

* The r.m.s. deviations from ideal geometry and r.m.s. variation in the B-factors of bonded atoms.

arm of the L, whereas the extended strands and β -turns form its base^{19–21} (Fig. 2b).

p19^{INK4d} binds through its concave face to the cleft between the N- and C-terminal domains of Cdk6, just to one side of the active site (Fig. 2). In p19^{INK4d}, the unusual ankyrin II^{19–21} is at the centre of the interface which mainly involves ankyrins I–III (Fig. 3a, b). In Cdk6, the interaction occurs mainly through the β -sheet in the N-terminal domain, loop L7 linking the N- and C-terminal domains, and helix α 2 in the C-terminal domain (Fig. 3a, c). As predicted¹⁹, the structure shows that INK4 protein binding does not obstruct the cyclin binding site, which is consistent with the formation of ternary complexes with cyclin D/Cdk6 (refs 11, 22) (D.H.B., S.W., L.B. and E.D.L., unpublished results).

Structure of the p19^{INK4d}/Cdk6 interface

The interface consists of three main parts. In the first, helices α 1, α 3 and α 5 of p19^{INK4d} pack almost at right angles across the N-terminal β -sheet of Cdk6 (Figs 2 and 4). In the second, the β -turns between ankyrins I/II and II/III in p19^{INK4d} provide a lid over the β -sheet of Cdk6. Together, these interactions form a clamp around the N-terminal β -sheet of Cdk6 (Fig. 2b). Finally, loops L4 and L6, between the helices in ankyrins II and III of p19^{INK4d}, respectively, clamp down on the C-terminal domain, moving the N-terminal domain relative to that of the C terminus. Loop L4 of p19^{INK4d} binds over the linker (Cdk6 loop L7), with the side chain of Met 50 to one side and Gly 52 to the other. Loop L6, in particular Phe 86, binds in a hydrophobic pocket between loops L7 and loop L11 in the C-

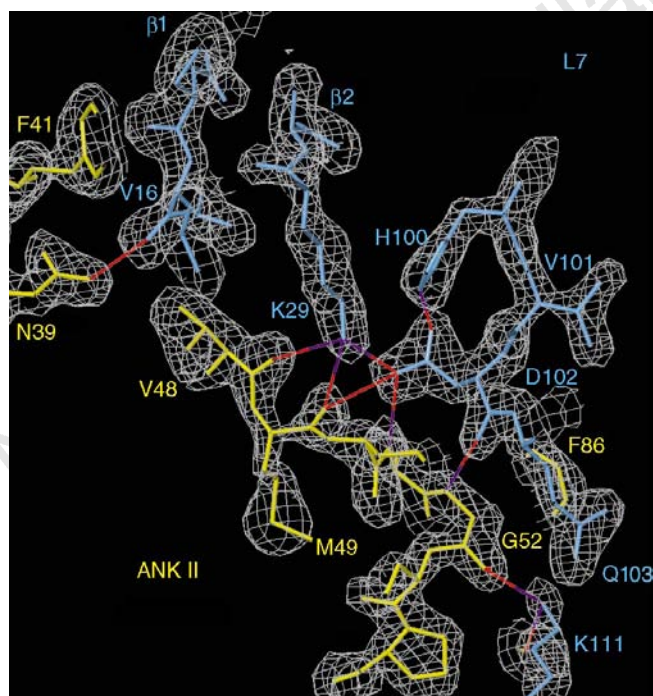


Figure 1 Electron density at the p19^{INK4d}/Cdk6 interface contoured at 1.2 σ . p19^{INK4d} and Cdk6 are coloured yellow and blue, respectively. Lys 29 and Asp 102, which form an ion pair in Cdk6, and residues in p19^{INK4d} involved in intermolecular hydrogen bonding, are labelled. Hydrogen bonds, in blue/red, are formed between: the carbonyl groups of Val 48 and Met 49 (p19^{INK4d}) and the side chain of Lys 29 (Cdk6); the carbonyl of Met 49 and amide of Phe 51 (p19^{INK4d}) with the side chain of Asp 102 (Cdk6); the amide of Gly 52 (p19^{INK4d}) with the carbonyl of Asp 102 (Cdk6); and the carbonyl of Gly 52 (p19^{INK4d}) with the side chain of Lys 111 (Cdk6). The (2|F_O| – |F_C|)-simulated annealing omit map, in grey, was calculated at 1.9 Å using phases from a model in which all atoms from residues 98 to 104 in Cdk6 were omitted together with all those within the surrounding 5 Å. The model was subjected to refinement by slow cooling from 2,500 K using X-PLOR⁴³. Figures 1 and 5c were produced using the computer program O⁴⁶.

terminal domain of Cdk6 (Figs 2a and 4). The total surface area buried in the complex is 1,700 Å² (Fig. 3a).

The interaction between p19^{INK4d} and Cdk6 partly buries several intermolecular ion pairs (Fig. 4). Apart from Arg 40 in p19^{INK4d} (Ser in p16^{INK4a} and Gly in p18^{INK4c}), appropriate positively or negatively charged groups are conserved between Cdk4 and Cdk6, but only Arg 31 is conserved between Cdk4/6 and Cdk2. In Cdk2, positions 14, 18 and 111 are Gln, Lys and Ala (Fig. 3c), which suggests that specificity of the INK4 proteins for Cdk4 and Cdk6 results in part from their inability to form complementary electrostatic surfaces with other CDKs. The unusual ankyrin II (refs 19–21) forms a network of intermolecular hydrogen bonds to the Cdk6 Lys 29–Asp 102 ion pair that is buried by the interaction (Figs 1 and 4). These interactions should also be conserved in Cdk4, but not in Cdk2, where His 100–Asp 102 is replaced by Phe 82–His 84 (Fig. 3c); replacement of residues 100–102 with those from Cdk2 abolishes the ability of p16^{INK4a} to inhibit the kinase (M.S., unpublished results).

His 119, the most C-terminal residue of p19^{INK4d} that interacts with Cdk6, packs to one side of loop L11 (Fig. 4a). This indicates that the shorter INK4 proteins (for example, p15^{INK4b}) will all make very similar interactions with Cdk4/Cdk6; the existence of active p16^{INK4a} mutants with C-terminal deletions supports this conclusion²³.

Two other structures of ankyrin repeat-domain complexes are known. 53BP2 interacts with p53 through the convex face of its L-shaped structure, on the other side of the β -turns to p19^{INK4d}/Cdk6 (ref. 24). The interaction between the α - and β -subunits in the GABA (γ -aminobutyric acid) α/β -DNA complex is more similar to that of p19^{INK4d}/Cdk6. The β -ankyrin repeat domain interacts with the α -subunit through its concave face, but this involves mainly the tips of the four β -turns in a more hydrophobic interface²⁵.

Analysis of mutations to p16^{INK4a}. The structure allows us to identify p16^{INK4a} mutations that affect interactions with Cdk4 and Cdk6, as opposed to destabilizing its structure^{19–21,26}. For example, four mutations found in p16^{INK4a} associated with inherited melanomas are equivalent to residues in p19^{INK4d} involved in the interaction with Cdk6: Gln 47 to Arg, Met 50 to Ile, Arg 83 to Pro and Asp 104 to Asn (p19 numbering; Figs 3a, b). Of these, Met 50 and Arg 83 become completely buried on complex formation (Fig. 3a); Met 50 packs alongside loop L7 and the beginning of helix α 2 in Cdk6, whereas Arg 83 tucks into a pocket near the top of Cdk6 strand β 3, making many interactions (Fig. 4a). The structure allows a complete analysis of all known mutations to p16^{INK4a}.

Analysis of mutations to Cdk4. Both of the known cancer-associated mutations in Cdk4 are to charged residues in the interface that are directly involved in p19^{INK4d}/Cdk6 interactions (see above and Fig. 3a, c). Mutation of Arg 24 in Cdk4 (Arg 31 in Cdk6) affects the ability of p16^{INK4a} to inhibit the kinase^{27,28} and is strongly linked to sporadic and hereditary melanomas^{27,29–31}. Mutation of Lys 22 to Ala in Cdk4 (Lys 29 in Cdk6) also affects the ability of p16^{INK4a} to inhibit the kinase²⁸ (M.S., unpublished results) and its mutation to Glu is linked to formation of sporadic melanomas³⁰.

Conformational changes in p19^{INK4d} and Cdk6

Conformational changes in p19^{INK4d}. The structure of p19^{INK4d} in solution and in the complex with Cdk6 can be superimposed with a root-mean-square (r.m.s.) deviation of 1.7 Å (over the C α carbons of residues 10–165). The difference is mainly due to a small long-range bending of the molecule and a small angular distortion in the packing of helices α 4 and α 6. The largest changes occur in the β -turns between ankyrins I/II, II/III and III/IV, which have few structural restraints in solution¹⁹, although they are not flexible (F.Y. Luh, S.J.A., P.J.D. and E.D.L., unpublished results).

Structure of Cdk6. In Fig. 3, the secondary structures, determined by crystallography, of Cdk2 (ref. 15) and Cdk6 in the p19^{INK4d} complex are compared; they are very similar, but Cdk6 (and Cdk4)

have two significant insertions (residues 70–72 and 86–90 in Cdk6). The X-ray analysis shows that the first binds a calcium ion, which is coordinated octahedrally by the carboxylates of Glu 69 and Glu 72 and four water molecules. This insertion extends the helix by one turn, which may link it more firmly to loop L5, which anchors the two domains of Cdk6 together (see below). The second insertion in Cdk6 extends loop L6 and the β -sheet, involving Cdk6 strands β 3, β 4 and β 5. This insertion allows more extensive hydrogen bonding with loop L4, which ties the N terminus of the PLSTIRE helix more firmly to its tip than is possible in Cdk2 (Figs 2 and 5). The mitogen-activated protein kinase Erk2 also has a five-residue insertion between strands β 4 and β 5 (ref. 32). In Erk2, there is no interaction between the equivalent two loops, which indicates that this interaction in Cdk6 might be the result of p19^{INK4d} binding.

Conformational changes in Cdk6. Protein kinases have a 'closed' and an 'open' conformation. In the latter, the N-terminal β -sheet moves away from the active site³³. Comparison with cAMP-dependent protein kinase A³⁴ shows that in this structure, Cdk6, like those of Cdk2 (refs 15–17), has a 'closed' conformation.

The C-terminal domain of Cdk6 can be superimposed on the inactive Cdk2 (ref. 15), or Cdk2 in its active cyclin A complexes^{16,17}, with an r.m.s. deviation of ~ 0.5 Å (over the 70 α -helical C α atoms).

The N-terminal β -sheets can also be superimposed with an r.m.s. deviation of ~ 0.4 Å (over the 18 β -sheet C α atoms). Thus we expect that Cdk2 and Cdk6 by themselves will have very similar structures. A comparison of Cdk6/p19^{INK4d} with Cdk2 (ref. 15) indicates that the N-terminal domain of Cdk6 undergoes extensive movement on binding p19^{INK4d}. Loop L5 connects the C terminus of the PLSTIRE helix α 1 to Cdk6 strand β 4 (Fig. 4c). In all of the CDK structures, this loop, together with the linker (loop L7), anchors the N-terminal domain to the C terminus, thus providing a frame of reference for discussion (Fig. 5a).

Compared with isolated Cdk2, the N-terminal domain in p19^{INK4d}/Cdk6 is rotated approximately in the plane of the β -sheet about these fixed points, so that the different strands of the β -sheet now lie in positions similar to those seen in the Cdk2/cyclin A complex. In Cdk6, the whole β -sheet is also translated by about 5 Å (Fig. 5b). In the absence of a structure of Cdk6 by itself, we cannot be sure that this movement of the N-terminal domain in Cdk6 is due to p19^{INK4d} binding; for example, p19 could stabilize an inactive conformation (see below) that is already present in the isolated Cdk6.

The N-terminal β -sheet moves as a whole, but this movement is not concerted with the movement of the N terminus of the PLSTIRE

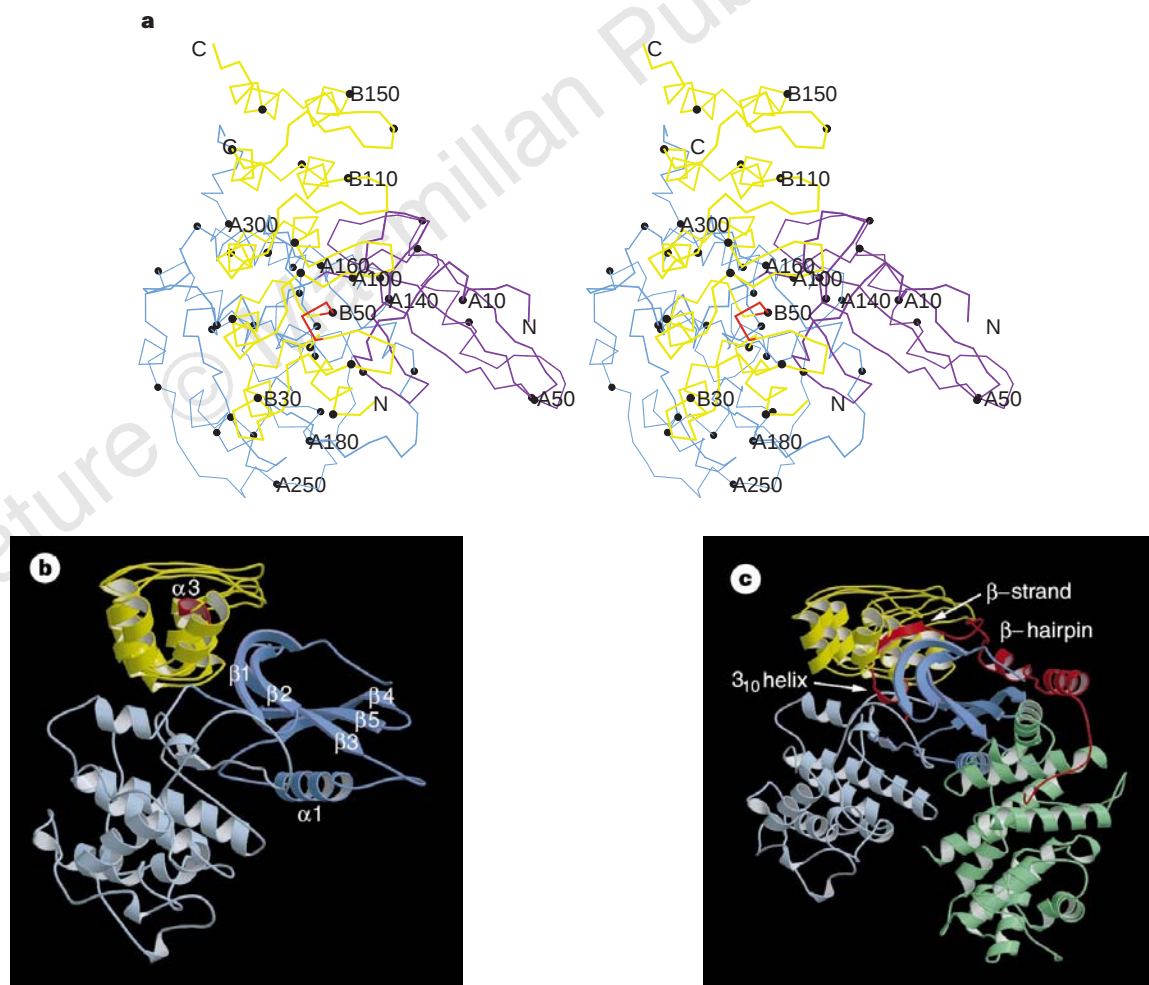


Figure 2 Three-dimensional structure of the p19^{INK4d}/Cdk6 complex. **a**, Stereo view of the C α trace of the p19^{INK4d}/Cdk6 structure. Every tenth C α atom is indicated by a solid ball. **b**, Schematic drawing of the same complex after rotation by 90° about the x axis. **c**, P19^{INK4d} and p27^{Kip1} prevent each other binding to the CDK subunit in cyclin D/Cdk6. The structure in **c** is rotated 20° about the z axis compared to **b**. In **a** and **b** p19^{INK4d} is coloured yellow, apart from helix α 3 (residues 46–50), which is red. The C-terminal domain of Cdk6 is coloured light blue,

whereas the N-terminal domain, which undergoes extensive movement, is dark blue. In **c**, p19^{INK4d} is coloured yellow, Cdk6 is light/dark blue, p27^{Kip1} is red and cyclin A is green. p27^{Kip1} and cyclin A from the p27^{Kip1}/cyclin A/Cdk2 structure⁴⁰ were superimposed, as described in Fig. 5a, on p19^{INK4d}/Cdk6 (r.m.s. deviation was 0.46 Å over 70 residues). Figures 2 and 5a, b were produced using the programs MOLSCRIPT⁴⁷ and RASTER3D⁴⁸.

helix $\alpha 1$, which is tethered by interactions between loops L4 and L6 (see above). Compared with the Cdk2 protein, the N terminus of the PLSTIRE helix $\alpha 1$ is pulled away by ~ 2.0 Å from the rest of the protein and it rotates about its axis by $\sim 30^\circ$ in a clockwise direction (viewed from the N terminus). This movement is in the opposite direction to that seen when cyclin A binds to Cdk2 (Fig. 5b), but the

rotation is in the same sense (this helix rotates $\sim 90^\circ$ in going from Cdk2 to Cdk2/cyclin A¹⁶).

Implications for catalysis

Cyclin-induced activation. Glu 61, in the PLSTIRE helix $\alpha 1$, is one of a triad of conserved catalytic residues (Lys 43, Glu 61 and Asp 163

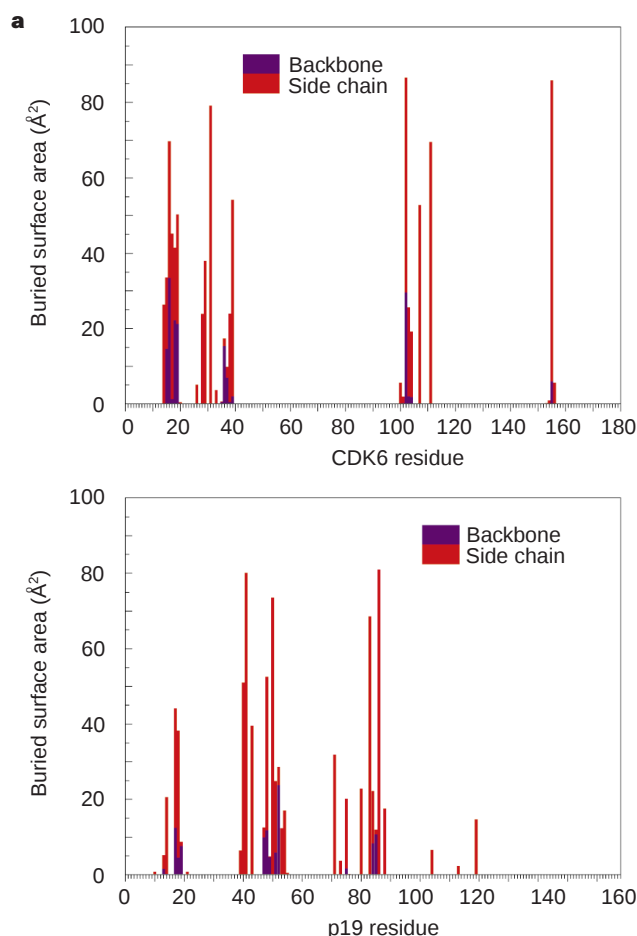
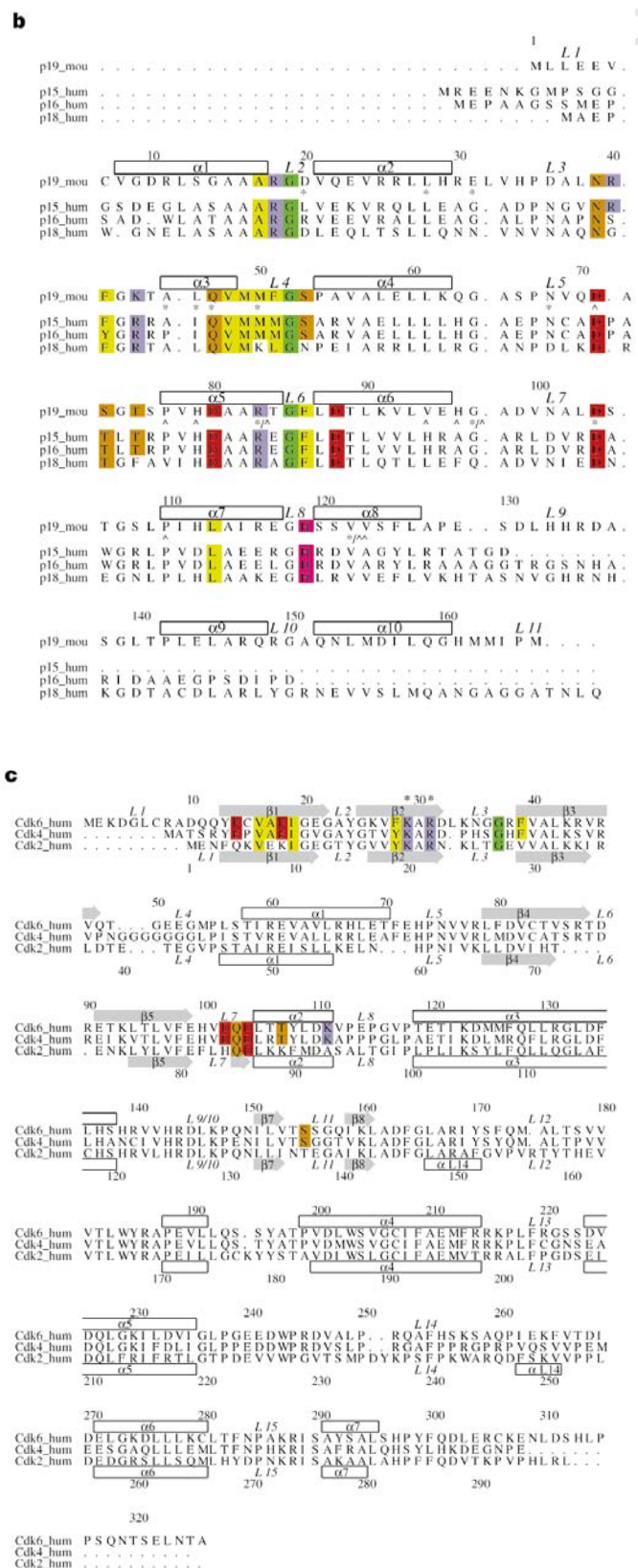


Figure 3 Sequence comparisons, secondary structure and residues buried in complex formation. **a**, Differences in solvent accessibility, calculated using the program NACCESS⁴⁹, for the isolated p19^{INK4d} and Cdk6 subunits compared with the complex. Buried backbone and side chains are shown in blue and red, respectively. The interface contains around 17 charged and 14 large hydrophobic residues whose solvent accessibility is substantially reduced on complex formation. **b**, Sequence of the mouse p19^{INK4d} protein showing the homology between the different ankyrin repeats and between the different members of the INK4 family. Positions in p16^{INK4a} that are mutated in familial melanomas are indicated by an asterisk⁵⁰; positions known to effect p16^{INK4a} binding or inhibition of Cdk4 are indicated by a circumflex symbol ($\hat{}$). **c**, Sequences of Cdk6, Cdk4 and Cdk2. In addition to those discussed in the text, there is an insertion at residue 33 and deletions at residues 175, 194 and 251–252 compared to Cdk2. Positions 22 and 24 in Cdk4, where mutations inhibit binding of p16^{INK4a} (ref. 28; M.S., unpublished results) and which are implicated in the formation of melanomas^{27,30,31} are indicated by an asterisk. In both **b** and **c**, conserved residues involved in the interaction are coloured yellow (hydrophobic), orange (asparagine, glutamine, serine and threonine), green (glycine), red (acidic), magenta (histidine) and blue (basic); arrows (grey) and rectangles (white) indicate the positions of β -strands and α -helices, respectively.



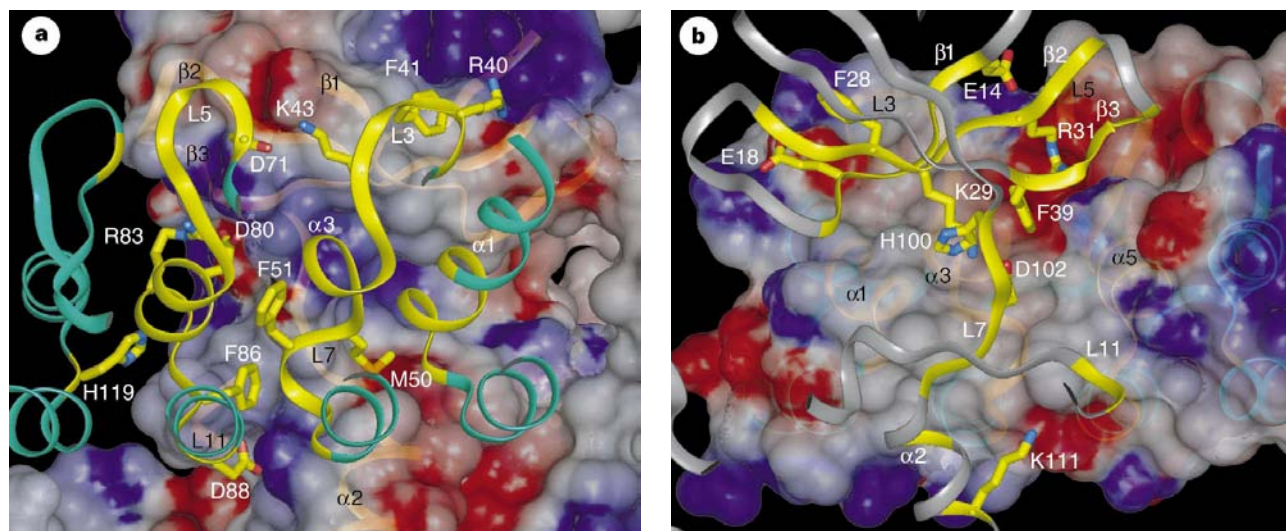


Figure 4 Surface representation of the p19^{INK4d}/Cdk6 interface. The electrostatic potential at either the p19^{INK4d} or Cdk6 protein surface is contoured and colour coded at -5.0 kT (red) and $+5.0$ kT (blue). The interacting regions of both p19^{INK4d} and Cdk6 are yellow (otherwise green (p19^{INK4d}) and grey (Cdk6)); labels for the interacting protein are in white, whereas those for the contoured protein are in black. Side chains of residues involved in electrostatic interactions are red (negative) and blue (positive). Electrostatic interactions are formed between: Arg 40 (p19^{INK4d}) and Glu 18 (Cdk6); Lys 43 (p19^{INK4d}) and Glu 14 (Cdk6); Asp 70 (p19^{INK4d}) and Arg 31 (Cdk6); Asp 88 (p19^{INK4d}) and Lys 111 (Cdk6). Hydrophobic aromatic interactions are formed between Phe 41 (p19^{INK4d}) and Phe 28 (Cdk6),

and between Phe 51 (p19^{INK4d}) and Phe 39 (Cdk6). Other p19^{INK4d}/Cdk6 interactions not mentioned in the text, or in Fig. 1, involve hydrogen bonds between the sidechains of Arg 31 (Cdk6) and Asp 71 (p19^{INK4d}) and between the side chain of Ser 155 (Cdk6) and the carbonyl groups of Thr 84 and Gly 85 (p19^{INK4d}). Two views of the structure are shown. **a**, Surface representation of Cdk6, in approximately the same orientation as in Fig. 2a, showing the interactions with p19^{INK4d}. **b**, Surface representation of p19^{INK4d}, after rotating the structure by 180°, showing the interactions with Cdk6. This figure was produced using Insight II (MSI) and significant portions of the structure were clipped for clarity.

in Cdk6) that are required for the correct orientation of ATP and coordination of magnesium in the active site of all eukaryotic kinases³³. In Cdk2, helix $\alpha 1$ and this glutamate move into the active site only on interaction with cyclin A^{16,17}. If p19^{INK4d} binding prevented this movement it would inhibit the enzyme. In Cdk2 and Cdk6/p19^{INK4d}, the active-site Glu residues are in a similar position, whereas it is over 10 Å away from the position seen in the active Cdk2/cyclin A structures, where it should interact with Lys 43 and coordinate the magnesium ion. Could cyclin binding move the PLSTIRE helix $\alpha 1$ and Glu 61 to their correct positions?

The similarity between the D-type cyclins and cyclin A suggests they will contact their respective CDKs in similar ways. For helix $\alpha 1$ to take up its active position, it must move in towards the C-terminal domain and loops L4 and L6 of Cdk6 must move out of the way of helices $\alpha 5$ and $\alpha 3$ in the cyclin¹⁶, respectively (Figs 2c and 5b). The interaction between loops L4 and L6 in p19^{INK4d}/Cdk6 might therefore inhibit the necessary movement of helix $\alpha 1$. In Cdk4, however, several glycine residues are inserted in loop L4 that are not present in Cdk6 or Cdk2 (Fig. 3c). This sequence may therefore be unstructured in Cdk4, which would not favour a close interaction between loops L4 and L6. However, even if the interaction between L4 and L6 in Cdk6 did break down on cyclin binding, helix $\alpha 1$ might still be prevented from moving to its active position by a clash with the tip of loop L2 (Fig. 5b). In conclusion, the structure suggests that p19^{INK4d} binding might prevent the cyclin-induced conformational change needed to activate the kinase. This conclusion is consistent with indications that p16^{INK4a} binding might weaken the affinity of cyclin D for Cdk4 and Cdk6 *in vivo*³⁵ and *in vitro*³⁶.

Conformation of the T-loop. In Cdk2, the T-loop interacts with strand $\beta 1$ and obstructs the ATP- and substrate-binding sites, as well as the threonine side chain, which must be phosphorylated by CAK¹⁵. The conformation of the T-loop in Cdk6 (residues 164–183) is similar to that in Cdk2. As with Cdk2, however, there is no reason why the T-loop in Cdk6 could not move out of the way to the position seen in the Cdk2/cyclin A structures^{16,17} (Fig. 5). In fact, residues 173–179 have poor electron density, indicating conforma-

tional flexibility in this loop. The conformation of this loop is very different to that seen in a second structure of p19^{INK4d}/Cdk6, which may be due to different crystallization conditions or different crystal packing³⁷. The two structures thus provide different views of the accessible conformations of this loop.

Implications for ATP binding. In Cdk2, ATP is bound by residues in the N-terminal domain (loop L2, strand $\beta 3$ including Lys 43, and Val 77 in strand $\beta 5$) by residues in the linker (loop L7), and by residues in the C-terminal domain (loop L10, strand $\beta 7$, and Ala 162/Asp 163)^{17,33,34} (all Cdk6 numbering; Fig. 3c). Residues in the linker L7 and in the C-terminal domain are in similar positions in p19^{INK4d}/Cdk6 and in all of the Cdk2 structures, whereas residues in the N-terminal domain are in different positions (Fig. 5c).

P19^{INK4d} does not block ATP binding directly, although the side chain of Arg 18 has an alternative conformation where ATP must bind. When p19^{INK4d} interacts with Lys 29 and Asp 102 of Cdk6, the side chain of His 100 in Cdk6 moves into the adenine pocket of the ATP-binding site, and is held in this position by hydrogen bonding between the imidazole NE2 nitrogen and the carboxylate of Asp 102 (Figs 1 and 5c). This positioning of His 100 provides one possible mechanism by which ATP binding might be altered.

In the N-terminal domain, hydrogen bonding between the amides in the phosphate-binding loop L2 (residues Ala 23, Tyr 24 and Gly 25) and the β -phosphate of ATP is required to position the phosphates during catalysis. The side chain of Val 27 should also interact with the ribose and adenine rings^{17,34}. Modelling of ATP into the p19^{INK4d}/Cdk6 structure suggests that the movement of Cdk6 loop L2, relative to the other regions of the ATP-binding site, would prevent this hydrogen bonding and therefore prevent productive interaction with ATP (Fig. 5c).

Competition between CDKIs

The Cip/Kip and INK4 families of CDKIs compete for binding to D-type cyclin/CDKs^{22,38,39}. In the p27^{Kip1}/Cdk2/cyclin A complex, a β -hairpin in p27^{Kip1} interacts with the N-terminal β -sheet of Cdk2 and a β -strand in p27^{Kip1} displaces strand $\beta 1$ of Cdk2. In addition, a

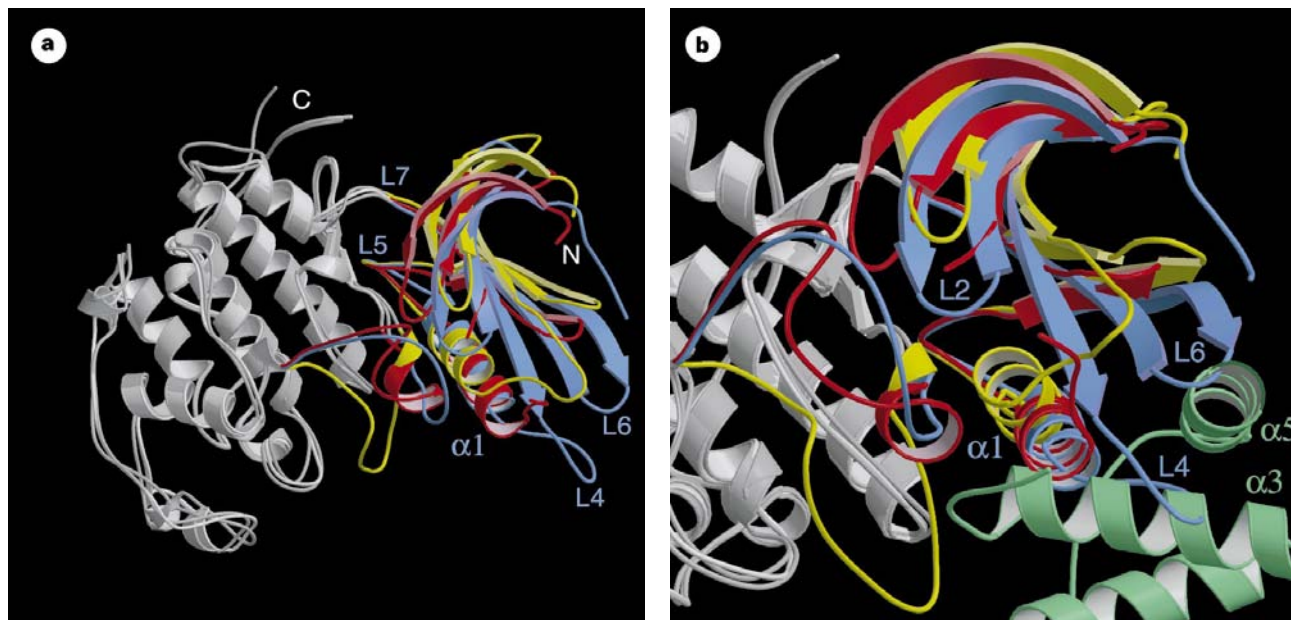


Figure 5 Comparison of ATP-binding sites in different kinases. **a**, Comparison of the structure of p19^{INK4d}/Cdk6 with those of Cdk2 (ref. 15) and Cdk2 from the cyclin A complex^{16,17}. Cdk6 was superimposed on Cdk2 (r.m.s. deviation was 0.56 Å over 70 residues) and Cdk2 from the cyclin A complex (r.m.s. deviation was 0.49 Å over 70 residues) by aligning the Cα atoms in the α-helices in the C-terminal domain of both proteins. Loop L5 and the linker (loop L7) (Fig. 3c), which anchor the N- and C-terminal domains together, are labelled. **b**, The same structures, showing only the N-terminal domain, illustrating the changes in position of the N-terminal β-sheet and helix α1. In both **a** and **b**, the N-terminal domains and T-loop are coloured red (Cdk2), yellow (Cdk2 from the cyclin A complex) and blue (Cdk6); other regions of all three proteins are coloured grey. In **b**, helices α3 and α5 of cyclin A, which interact with PSTAIRE helix α1 in cyclin A/Cdk2 (refs 16, 17), are green. **c**, The ATP-binding site in Cdk6. Lys29, His 100 and Asp 102, which together might inhibit ATP binding, are labelled, as are key active-site residues and the phosphate-binding loop L2. ATP, from the active Cdk2/cyclin A complex¹⁷, is superimposed on the structure with the N1 and N6 adenine nitrogens making conserved hydrogen bonds with the carbonyl of Glu99 and amide of Val101, respectively^{17,34}. The structures of Cdk6 and ATP are yellow and green, respectively. Carbon, nitrogen, oxygen and phosphorus atoms are yellow, blue, red and yellow, respectively. The structure of Cdk2 from the cyclin A/Cdk2 complex¹⁷ is blue.

p27^{Kip1} 3₁₀ helix binds in the catalytic cleft⁴⁰. As with Cdk2/cyclin A⁴⁰, p27^{Kip1} interacts with both the CDK and the cyclin components of Cdk4/cyclin D and Cdk6/cyclin D⁴¹. Comparison of the p19^{INK4d}/Cdk6 and p27^{Kip1}/Cdk2/cyclin A structures shows that, although the interaction of the p27 β-hairpin might still be possible, the other p27 interactions would be inhibited by an INK4 protein. In other words, binding of one class of inhibitor should preclude the binding of the other (Fig. 2c). This is supported by experiments that show that either p15^{INK4b} (ref. 22), p16^{INK4a} or p19^{INK4d} (D.B., S.W., X.X., L.B. and E.D.L., unpublished results) prevent binding of p27^{Kip1} to the kinase subunit in cyclin D/CDKs and vice versa.

Conclusions

The structure of the p19^{INK4d}/Cdk6 complex allows us to understand CDK regulation by the INK4 family of CDKIs. Close interactions between p19^{INK4d} and loop L7 of Cdk6 are likely to be key determinants of specificity. The structure indicates that the interaction results in extensive movement of the N-terminal domain of

the kinase, relative to the C terminus, inhibiting productive binding with ATP and movement of the PLSTIRE helix to its active conformation. It is possible that conformational rearrangements could allow movement of this helix to its active conformation on cyclin binding, but productive ATP binding would be unlikely. The structure shows how the Cip/Kip and INK4 families of CDKI prevent each other from binding to the CDK, allowing us to understand how they coordinate regulation of the cell cycle. □

Methods

Protein expression and purification. Mouse p19^{INK4d} was expressed in *Escherichia coli* and purified as described¹⁹. Human Cdk6 was expressed in SF9 insect cells infected with a glutathione S-transferase (GST)–Cdk6 baculovirus. After infection, the cells were collected after 70 h. Excess p19^{INK4d} protein was added before sonication and removal of cell debris by ultracentrifugation. The complex was purified on glutathione–sepharose resin (Pharmacia) and, after elution with 25 mM glutathione, cleaved from GST by overnight digestion with thrombin. After treatment with antithrombin III resin, the GST was removed

with glutathione-Sepharose resin and the complex was further purified on Mono-Q and Sephadex S-200 columns (Pharmacia). The protein (>95% pure) was concentrated to ~25 mg ml⁻¹ in 10 mM Tris, pH 7.5, 350 mM NaCl, 10 mM DTT and 10% glycerol. Electrospray mass spectrometry confirmed the identity and molecular mass of both proteins.

Crystallization and data collection. Crystals were grown at 4°C by the hanging-drop vapour-diffusion method, after mixing the complex with an equal volume of buffer containing 6% PEG 8000, 0.1 M MES, pH 6.5, and 0.1 M calcium acetate. The crystals form in space group $P2_12_12_1$ with $a = 74.21$, $b = 76.41$ and $c = 93.80$ Å and contain one molecule of the complex in the asymmetric unit. Diffraction data were collected at synchrotron sources at room temperature using either a large MAR image-plate detector at beamline Px 9.5 (Daresbury Laboratory) or a CCD detector at beamline ID14 EH3 (ESRF Laboratory).

Molecular replacement. The structure was determined by molecular replacement (MR). The positions of the Cdk6 and p19^{INK4d} were determined by MR with the program AMORE⁴², using the structures of Cdk2 and p19 (PDB codes, 1hck and lap7, respectively). The rotation and translation functions for Cdk6 were determined first and then fixed, before determining the translation function for p19^{INK4d}. The correlation coefficient was 40.9 for 10–4 Å data. The MR solution was consistent with results of site-directed mutagenesis of both p19^{INK4d} and Cdk4, which showed that mutation of Phe 86 in p19 abolished binding to Cdk6 (data not shown) and that mutation to positions 22, 24 and 95–97 in Cdk4 abolished binding of p16^{INK4a} (M.S., unpublished results). In addition, it was consistent with two-dimensional ¹H–¹⁵N HSQC NMR spectra that show that the amides of residues His 29, Arg 40, Gly 42, Thr 44, Ala 45, Ser 66 and Ser 76 in p19^{INK4d} have large changes in chemical shift, whereas those of Asp 10, Ser 13, Arg 18, Gly 19, Met 50, Phe 51, Gly 52 and Gly 85 have smaller shifts, on complex formation. The structure was refined using torsion-angle dynamics in X-PLOR⁴³, followed by maximum-likelihood refinement using REFMAC⁴⁴. After successive rounds of rebuilding into SIGMAA-weighted maps⁴⁵ using O (ref. 46), the final model consists of residues 5–309 of Cdk6, residues 6–165 of p19^{INK4d}, one calcium ion and 286 water molecules.

Received 15 May; accepted 20 July 1998.

- Sherr, C. J. Cancer cell cycles. *Science* **274**, 1672–1677 (1996).
- Morgan, D. O. Principles of CDK regulation. *Nature* **374**, 131–134 (1995).
- Parker, L. L., Sylvestre, P. J., Brynes, M. J., Liu, F. & Piwnicaworms, H. Identification of a 95 Kda WEE1-like tyrosine kinase in Hela-cells. *Proc. Natl Acad. Sci. USA* **92**, 9638–9642 (1995).
- Kumagai, A. & Dunphy, W. G. The Cdc25 protein controls tyrosine dephosphorylation of the Cdc2 protein in a cell-free system. *Cell* **64**, 903–914 (1991).
- Sherr, C. & Roberts, J. Inhibitors of mammalian G1 cyclin-dependent kinases. *Genes Dev.* **9**, 1149–1153 (1995).
- Harper, J. W. in *Checkpoint Control and Cancer* (ed Kastan, M. B.) Vol. 29, 91–107 (CSHL Press, New York, 1997).
- Soos, T. J. et al. Formation of p27-CDK complexes during the human mitotic cell cycle. *Cell Growth Differ.* **7**, 135–146 (1996).
- Poon, R. Y. C., Toyoshima, H. & Hunter, T. Redistribution of the CDK inhibitor p27 between different cyclin-CDK complexes in the mouse fibroblast cell cycle and in cells arrested with lovastatin or ultraviolet irradiation. *Mol. Biol. Cell* **6**, 1197–1213 (1995).
- Serrano, M., Hannon, G. J. & Beach, D. A new regulatory motif in cell-cycle control causing specific inhibition of cyclin D/CDK4. *Nature* **366**, 704–707 (1993).
- Hannon, G. J. & Beach, D. p15^{INK4B} is a potential effector of TGF-β-induced cell cycle arrest. *Nature* **371**, 257–261 (1994).
- Hirai, H., Roussel, M. F., Kato, J.-Y., Ashmun, R. A. & Sherr, C. J. Novel INK4 proteins, p19 and p18, are specific inhibitors of the cyclin D-dependent kinases CDK4 and CDK6. *Mol. Cell. Biol.* **15**, 2672–2681 (1995).
- Chan, F. K. M., Zhang, J., Cheng, L., Shaprio, D. N. & Winoto, D. A. Identification of human and mouse p19, a novel CDK4 and CDK6 inhibitor with homology to p16^{INK4}. *Mol. Cell. Biol.* **15**, 2682–2688 (1995).
- Guan, K. et al. Isolation and characterization of p19^{INK4d}, a p16 related inhibitor specific to CDK6 and CDK4. *Mol. Biol. Cell* **7**, 57–70 (1996).
- Weinberg, R. A. The retinoblastoma protein and cell cycle control. *Cell* **81**, 323–330 (1995).
- De Bondt, H. L. et al. Crystal structure of cyclin-dependent kinase 2. *Nature* **363**, 595–602 (1993).
- Jeffrey, P. D. et al. Mechanism of CDK activation revealed by the structure of a cyclin-CDK2 complex. *Nature* **376**, 313–320 (1995).
- Russo, A. A., Jeffrey, P. D. & Pavletich, N. P. Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nature Struct. Biol.* **3**, 696–700 (1996).

- Bork, P. Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally. *Proteins Struct. Funct. and Genet.* **17**, 363–374 (1993).
- Luh, F. Y. et al. Structure of the cyclin-dependent kinase inhibitor p19^{INK4d}. *Nature* **389**, 999–1003 (1997).
- Venkataramani, R., Swaminathan, K. & Marmorstein, R. Crystal structure of the CDK4/6 inhibitory protein p18^{INK4} provides insights into ankyrin-like repeat structure/function and tumor-derived p16^{INK4} mutations. *Nature Struct. Biol.* **5**, 74–81 (1998).
- Byeon, L.-J. L. et al. Tumor suppressor p16^{INK4a}: Determination of solution structure and analyses of its interaction with cyclin-dependent kinase 4. *Mol. Cell* **1**, 421–431 (1998).
- Reynisdottir, I. & Massagué, J. The subcellular locations of p15^{INK4b} and p27^{Kip1} coordinate their inhibitory interactions with Cdk4 and Cdk2. *Genes Dev.* **11**, 492–503 (1997).
- Yang, R., Gombart, A. F., Serrano, M. & Koeffler, H. P. Mutational effects on the p16^{INK4a} tumour-suppressor protein. *Cancer Res.* **55**, 2503–2506 (1995).
- Gorina, S. & Pavletich, N. P. Structure of the p53 tumor suppressor bound to the ankyrin and SH3 domains of 53BP2. *Science* **274**, 1001–1005 (1996).
- Batchelor, A. H., Piper, D. E., de la Brousse, F. C., McKnight, S. L. & Wolberger, C. The structure of GABPα/β: an ETS domain-ankyrin repeat heterodimer bound to DNA. *Science* **279**, 1037–1041 (1998).
- Zhang, B. & Peng, Z.-Y. Defective folding of mutant p16^{INK4a} proteins encoded by tumor-derived alleles. *J. Biol. Chem.* **271**, 28734–28737 (1996).
- Wölfel, T. et al. A p16^{INK4a}-insensitive CDK4 mutant targeted by cytolytic T-lymphocytes in a human-melanoma. *Science* **269**, 1281–1284 (1995).
- Coleman, K. G. et al. Identification of Cdk4 sequences involved in Cyclin D1 and p16 binding. *J. Biol. Chem.* **272**, 18869–18874 (1997).
- Zuo, L. et al. Germline mutations in the p16^{INK4a} binding domain of CDK4 in familial melanoma. *Nature Genet.* **12**, 97–99 (1996).
- Tsao, H., Benoit, E., Sober, A. J., Thiele, C. & Haluska, F. G. Novel mutation in the p16/CDKN2A binding region of cyclin-dependent kinase-4 gene. *Cancer Res.* **58**, 109–113 (1998).
- Soufir, N. et al. Prevalence of p16 and CDK4 germline mutations in 48 melanoma-prone families in France. *Hum. Mol. Genet.* **7**, 209–216 (1998).
- Zhang, F., Strand, A., Robbins, D., Cobb, M. H. & Goldsmith, E. J. Atomic structure of the MAP kinase ERK2 at 2.3 Å resolution. *Nature* **367**, 704–710 (1994).
- Taylor, S. S. & Radzio-Andzelm, E. Three protein kinase structures define a common motif. *Structure* **2**, 345–355 (1994).
- Zheng, J. H. et al. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* **32**, 2154–2161 (1993).
- Parry, D., Bates, S., Mann, D. J. & Peters, G. Lack of cyclin D-Cdk complexes in Rb-negative cells correlates with high levels of p16^{INK4a/MTS1} tumour suppressor gene product. *EMBO J.* **14**, 503–511 (1995).
- DellaRagione, F. et al. Biochemical characterisation of p16^{INK4}- and p18-containing complexes in human cell lines. *J. Biol. Chem.* **271**, 15942–15949 (1996).
- Russo, A. A., Tong, L., Lee, J.-O., Jeffrey, P. D. & Pavletich, N. P. Structural basis for inhibition of the cyclin-dependent kinase Cdk6 by the tumour suppressor p16^{INK4a}. *Nature* **395**, 237–244 (1998).
- Reynisdottir, I., Polyak, A., Iavarone, A. & Massagué, J. Kip/Cip and INK4D CDK inhibitor cooperate to induce cell cycle arrest in response to TGF-β. *Genes Dev.* **9**, 1831–1845 (1995).
- Sandhu, C. et al. TGFβ stabilizes p15^{INK4b} protein, increases p15^{INK4b}/CDK4 complexes, and inhibits cyclin D/CDK4 association in human mammary epithelial cells. *Mol. Cell. Biol.* **17**, 2458–2467 (1997).
- Russo, A. A., Jeffrey, P. D., Patten, A. K., Massagué, J. & Pavletich, N. P. Crystal structure of the p27^{Kip1} cyclin-dependent kinase inhibitor bound to the cyclin A-Cdk2 complex. *Nature* **382**, 325–331 (1996).
- Vlach, J., Hennecke, S. & Amati, B. Phosphorylation-dependent degradation of the cyclin-dependent kinase inhibitor p27^{Kip1}. *EMBO J.* **16**, 5334–5344 (1997).
- Collaborative computing project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
- Brünger, A. T. X-PLOR, Version 3.1. A System for X-Ray Crystallography and NMR (Yale Univ. Press, New Haven and London, 1992).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Read, R. J. Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Crystallogr. A* **42**, 140–149 (1986).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron-density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Merrit, E. A. & Murphy, M. E. Raster3D. A program for photorealistic molecular graphics. *Acta Crystallogr. D* **50**, 869–873 (1994).
- Hubbard, S. J. & Thornton, J. M. 'NACCESS' computer program (Department of Biochemistry and Molecular Biology, Univ. College London, 1993).
- Foulkes, W. D., Flanders, T. Y., Pollock, P. M. & Hayward, N. K. The CDKN2A (p16) gene and human cancer. *Mol. Med.* **3**, 5–20 (1997).

Acknowledgements. We thank C. J. Sherr (Howard Hughes Medical Institute) for the plasmid expressing GST–p19^{INK4d}; A. Raine for computer programming; M. Nilges for X-PLOR scripts; I. Tickle for advice; F. Luh for the selectively protonated p16^{INK4d}; J. Eckstein for help with molecular graphics; and E. Duke, W. Burmeister and R. C. Kehoe for help with data collection. This work was supported by grants from the BBSRC and the Wellcome Trust. D.H.B. is supported by a studentship from the MRC. M.S. is supported by a grant from DGICYT, Spain, and by a core grant to the Department of Immunology and Oncology from Pharmacia-Upjohn and the Spanish Research Council. The Cambridge Centre for Molecular Recognition is supported by the BBSRC and the Wellcome Trust.

Correspondence and requests for materials should be addressed to E.D.L. (e-mail: e.d.laue@bioc.cam.ac.uk). The coordinates have been deposited in the Brookhaven Protein Data Bank (accession number 1blx).

Imaging circumstellar environments with a nulling interferometer

Philip M. Hinz*, J. Roger P. Angel*, William F. Hoffmann*, Donald W. McCarthy Jr*, Patrick C. McGuire*, Matt Cheselka*, Joseph L. Hora† & Neville J. Woolf*

* Steward Observatory, 933 N. Cherry Avenue, Tucson, Arizona 85721, USA

† Smithsonian Astrophysical Observatory, 60 Garden Street, Cambridge, Massachusetts 02138, USA

Extrasolar planets must be imaged directly if their nature is to be better understood. But this will be difficult, as the bright light from the parent star (or rather its diffracted halo in the imaging apparatus) can easily overwhelm nearby faint sources. Bracewell has proposed¹ a way of selectively removing starlight before detection, by superposing the light from two telescopes so that the stellar wavefronts interfere destructively. Such a 'nulling' interferometer could be used in space to search for extrasolar Earth-like planets through their thermal emission and to determine through spectroscopic analysis if they possess the atmospheric signatures of life²⁻⁴. Here we report mid-infrared observations using two co-mounted telescopes of the Multiple Mirror Telescope that demonstrate the viability of this technique. Images of unresolved stars are seen to disappear almost completely, while light from a nearby source as close as 0.2 arcsec remains, as shown by images of Betelgeuse. With this star cancelled, there remains the thermal image of its surrounding, small dust nebula. In the future, larger ground-based interferometers that correct for atmospheric distortions (using adaptive optics) should achieve better cancellation, allowing direct detection of warm, Jupiter-size planets and faint zodiacal dust around other nearby stars⁵.

With a conventional telescope, the very high contrast ratio needed to resolve a planet from a star requires diffraction-limited resolution several times sharper than the angular separation. At the longer thermal wavelengths, of most interest for imaging warm planets, telescope apertures of tens of metres would be needed. In Bracewell's interferometric method two small apertures suffice, provided that the stellar wavefronts are exactly superposed out of phase with no relative tilt, so that the cancellation is in detail across the whole pupil. Then the transmission pattern on the sky, for monochromatic light, is one of finely spaced fringes given by:

$$T(\theta) = \sin^2 \frac{\pi \theta d}{\lambda} \quad (1)$$

where d is the element spacing, λ is the wavelength of observation, and θ is the angular distance from the line on the sky through the star perpendicular to the interferometer baseline. Our test of Bracewell's concept at the Multiple Mirror Telescope used two telescopes spaced by 5 m with detection at 10- μ m wavelength. Then the fringe spacing on the sky is 0.4 arcsec, and full constructive interference for a source is only 0.2 arcsec from the nulled star (equation (1)).

The optical configuration (Fig. 1), is based on previously proposed designs with a single pass beamsplitter^{6,7} for a single nulled output. The wavefronts are translated for superposition without relative rotation or tilt. In operation, the 10- μ m image is observed live on a computer monitor while the beamsplitter is translated to adjust path length. The star intensity remains steady until the path lengths are nearly equal, when it is seen to flicker. This is caused by atmospheric turbulence which induces path-length changes of $\pm 5 \mu$ m, enough to shift the interference randomly between constructive and destructive states relative to its non-interfered flux. The minima are deepest when the average phase difference between

the wavefronts is 180° and the average slopes are the same (that is, the individual images are coincident).

Figure 2a shows the brightest and faintest snapshot images of the star α Tauri, which is unresolved at our baseline. The ability of the interferometer to cause the entire stellar Airy pattern to disappear is clearly shown. The right-hand nulled image has a peak intensity 4.0% and a total integrated flux of 6.0% that of the left-hand constructive image. This small residue arises because phase and tilt differences between the beams are not ideal or stable through even the best nulled exposure, and because higher-order atmospheric aberrations remained uncorrected. But already from this single, short exposure we can deduce that there are no companions lying in the constructive fringes as close as 0.2 arcsec from the star, to a level 3.5 magnitudes (25 times) fainter than the star. Weather did not allow observations through the night for rotated baseline orientation, as prescribed by Bracewell; combination and comparison of such data would have allowed the detection of any companion more than 1% of the star's brightness with separation ≥ 0.2 arcsec.

The ability of the interferometer to directly reveal emission sources close to a nulled star is illustrated by images of late-type giant stars α Orionis (Fig. 2b) and R Leonis (a single-frame image of R Leo is not shown). The faintest nulled images are still 36% and 24%, respectively, of the integrated flux of the brightest image. The

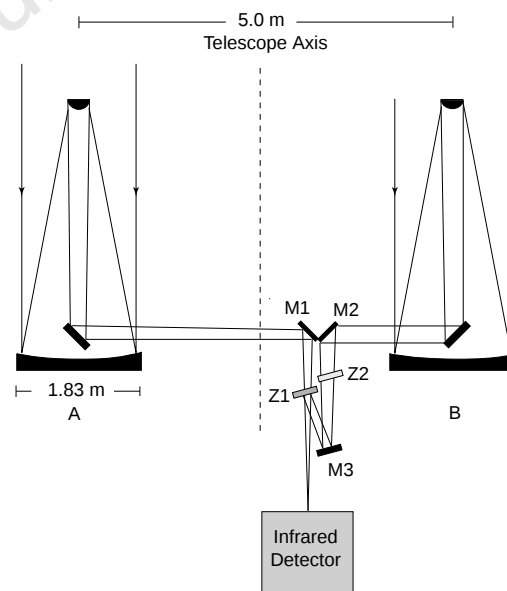


Figure 1 Schematic of our nulling interferometer at the MMT. The telescopes are co-mounted on a rigid frame pointed at the star, with no need for the variable delay lines for path equalization found in interferometers using separately mounted telescopes. The small number of reflections results in low thermal emission. Unmatched reflections are made at nearly normal incidence to minimize polarization differences. The $f/31$ folded Cassegrain beams are combined at the zinc selenide beamsplitter Z1, coated for equal transmission and reflection in the 8–13 μ m band. Beam A, after crossing the axis of the telescope, is folded down at M1, and is transmitted by the beamsplitter before coming to a focus. Beam B is folded downwards at M2 before reaching the axis of the telescope, through zinc selenide plate Z2 and back up at M3 to equalize the path lengths before being reflected from the underside of the beamsplitter. Achromatic 180° phase difference is realized by balancing a slight difference in air path with a 42- μ m path difference between the two zinc selenide elements, obtained by slight rotation of Z2. The residual chromatic error results in transmission ideally 0.008% over the band 8–12 μ m. In practice, the achromatic null is found initially by translating the beamsplitter vertically, with occasional subsequent small adjustments to correct for flexure. Image detection was made with the Rockwell (now Boeing) 128 $\times$ 128 arsenic-doped silicon BIB array of the Mid-infrared Array Camera, MIRAC2⁷. An internal cold stop limited the effective apertures of the telescopes to 1.6 m, for a diffraction-limited image width of 1.5 arcsec.

large residual fluxes arise because both stars have surrounding warm dust nebulae with diameters significantly more than 0.2 arcsec, and thus most of the dust nebulae lie outside the central null fringe. The stellar disks themselves are small enough (0.04 arcsec diameter) that only 0.4% of their flux is transmitted at the limb.

A feature in our implementation of the nulling interferometer different from Bracewell's original conception is that the focal plane shows an image of the field about the nulled star. The image formed is the object flux multiplied by the transmission pattern of equation (1) and convolved with the point-spread-function (PSF) of the individual elements. As the PSF is broader than the transmission pattern, the interference fringes are not visible in the focal plane. Thus true images are formed of smooth, extended features, such as circumstellar dust, whose structure is on a scale larger than the fringe spacing. α Ori and R Leo are shown in Fig. 3 with increased signal-to-noise in composite images, obtained from the 12 best nulled images by shifting to a common centre and adding, after re-pixelization. In the case of R Leo, the residual image with 1.7 arcsec full-width at half-maximum (FWHM) is barely wider than the

measured constructive image width of 1.5 arcsec, indicating a nebular width ≤ 1 arcsec. But Betelgeuse's dust nebula is clearly resolved, showing a FWHM of 2.4 arcsec and significant asymmetry. At 1% of the stellar intensity it extends radially 4 arcsec from the star. Our image with the star interferometrically removed is, to our knowledge, the first to show the nebula directly. Previous work has established the amount of dust and its spatial extent from spectral analysis and interferometric synthesis. Thus Danchi *et al.*⁸ have created visibility curves for α Ori and R Leo which have values of $\sim 60\%$ and $\sim 70\%$, respectively, for a 5-m baseline at $11\ \mu\text{m}$. Calculating our visibility in the same manner, using α Tau as our calibration for unit visibility, we obtain good agreement: $53 \pm 4\%$ for α Ori and $69 \pm 5\%$ for R Leo.

Figure 4 shows difference images formed by subtracting destructive images, each formed by recentering and co-adding the best individual exposures. The result for R Leo (Fig. 4a) shows a well defined diffraction pattern, out to the third Airy ring which is accentuated to 0.4% of the peak by the central obscuration of 10% of the diameter. When a difference image is formed for α Ori, a skewed Airy pattern results; however, by introducing a shift of 0.3 arcsec between the images, the nearly symmetric pattern shown (Fig. 4b) is recovered. The positions of the subtracted stars are shown by the crosses in Fig. 3, as determined by the offsets used to produce the symmetric patterns of Fig. 4. The offset for α Ori does not seem to be an effect of our instrument since it tracks field orientation in different exposures. We thus conclude that the centre of gravity of the nebular emission is displaced from the star. The offset angle is $\sim 25^\circ$ east of north. Asymmetries in the photosphere and immediate surroundings of α Ori have been observed in $\text{H}\alpha$

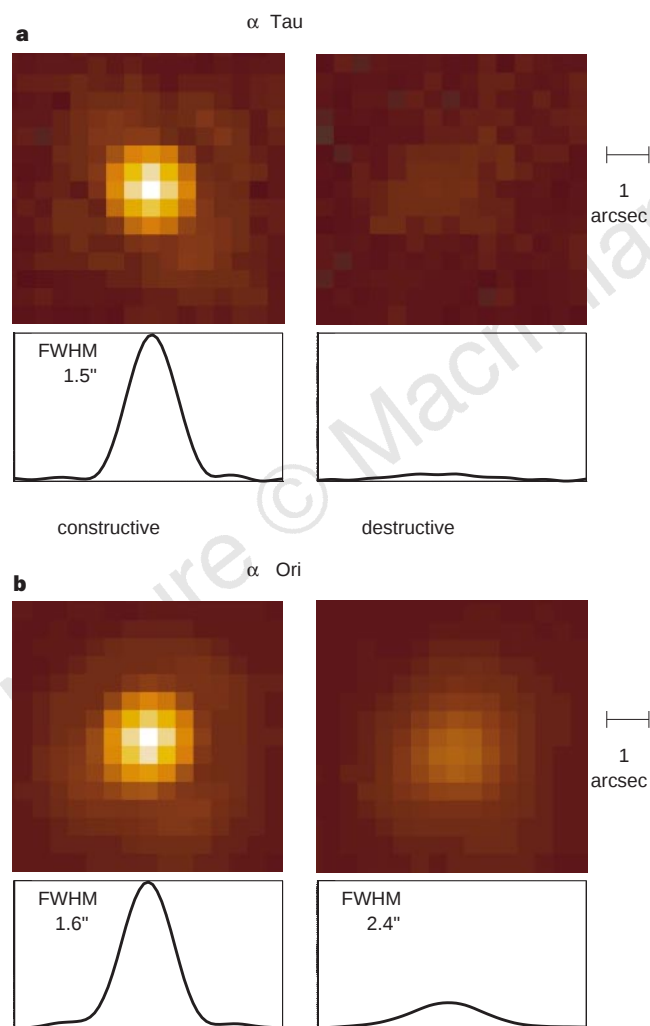


Figure 2 Single short-exposure frames at $10.3\text{-}\mu\text{m}$ wavelength, showing constructive and destructive interference. The second image of α Ori with the star nulled shows directly its surrounding dust nebula. The plots under the images are smoothed sections through the central row of pixels. These images were selected for maximal and minimal flux from a series of 1,000 50-ms frames of α Tau and 500 100-ms exposures of α Ori, taken at 10% bandwidth by MIRAC2 operating in a continuous recording mode with no "dead time". α Tau, which is unresolved, was observed on 17 February 1998 under seeing conditions characterized by Fried's length $r_0(10\ \mu\text{m}) = 3.7\text{ m}$. α Ori was observed on 17–18 January 1998, under better seeing conditions ($r_0 = 4.4\text{ m}$, slower motion).

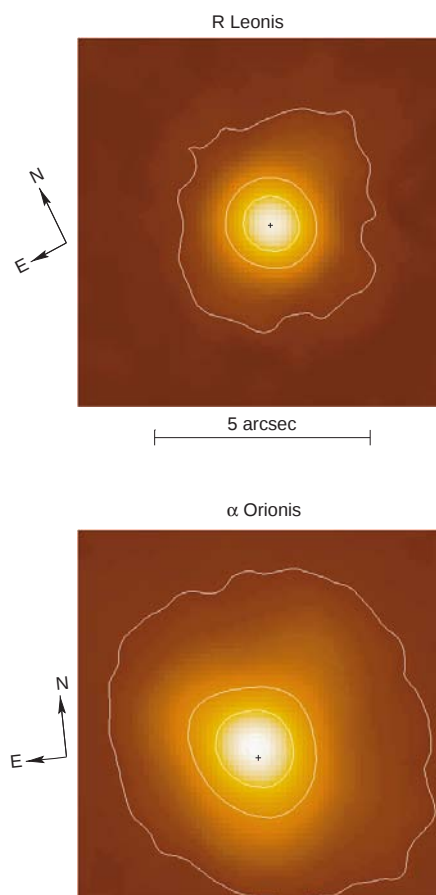


Figure 3 Averages of nulled images for α Ori and R Leo (also observed on 17–18 January 1998) showing the $10\text{-}\mu\text{m}$ dust nebulae with improved signal-to-noise ratio. The contours are at the level 1%, 10% and 20% of the non-interfered stellar peak intensity. The small + marks the centroid of the stellar emission. For α Ori, this is clearly offset from the nebula.

emission⁹, the ultraviolet continuum¹⁰ and 7-mm radio observations¹¹. Bester *et al.*¹² speculate the component of 10- μ m emission that they saw from newly forming dust in 1994 was probably asymmetric. However, the offset in centre of gravity of the extended 10- μ m emission has not been apparent in previous observations which were limited in their baseline orientation.

In the present system the best-nulled short exposures simply reflect the best destructive interference found among many images formed from atmospherically perturbed wavefronts. Change in phase and differential image motion over the integration time of the exposure prevent perfect starlight cancellation. From the measured rate of phase change, we expect these effects alone to result in a residual stellar intensity of 3% for α Tau¹³. Uncorrected higher-order errors across each aperture, such as astigmatism and focus, also contribute to the residual intensity. Based on Noll's analysis¹⁴ a differential phase error $\phi_{\text{rms}} = 0.5(d/r_0)^{5/6}$ is expected, adding a further 2% transmission of α Tau (Fried's length, $r_0 = 3.7$ m), for a total residual flux of 5%, in agreement with the measured value of 6%.

Future systems with phase and wavefront errors corrected in real time by adaptive optics should largely eliminate such residual

transmission. Correction of path length to 10 nm r.m.s. and high-order phase differences to 100 nm are projected, for star nulling at the 10^{-4} level. This development cannot be undertaken with the present instrument, as the MMT has now been dismantled, but a similar and much more powerful nulling configuration is planned for the Large Binocular Telescope (LBT), with 8.4-m primary mirrors also on a common mount¹⁵. With still a relatively short baseline (14.4 m centre to centre), the null fringe should be broad enough to suppress the ~ 1 mas disks of nearby stars by a factor of 10,000 at 10 μ m, so full advantage could be taken of the deep null of well-corrected wavefronts. Sensitivity to faint sources is determined by thermal background shot noise, which should be kept low in the LBT by making the adaptive correction at deformable secondary mirrors, with no additional optics. Thus there should be, as in Fig. 1, only three warm reflecting surfaces contributing to the background before the beam combination is made with cooled optics in a large cryostat. In this way, Bracewell interferometry from the ground should have the sensitivity, at 5- μ m wavelength, to find warm giant extrasolar planets to a 10 pc distance, to analyse them spectroscopically, and, at 10 μ m, to detect zodiacal dust clouds at solar level⁵. Zodiacal cloud measurements are crucial to the design of space systems to detect extrasolar Earth-like planets, because bright dust could overwhelm the much fainter planet signal¹⁶. $\square$

Received 23 April; accepted 26 June 1998.

1. Bracewell, R. N. Detecting nonsolar planets by spinning infrared interferometer. *Nature* **274**, 780–781 (1978).
2. Angel, J. R. P., Cheng, A. F. & Woolf, N. J. A space telescope for infrared spectroscopy of earthlike planets. *Nature* **322**, 341–343 (1986).
3. Leger, A. *et al.* Could we search for primitive life on extrasolar planets in the near future? *Icarus* **123**, 249–255 (1996).
4. Woolf, N. J. & Angel, J. R. P. in *Planets Beyond the Solar System and the Next Generation of Space Missions* (ed. Soderblom, D.) 285–293 (ASP Conf. Ser. 119, Astron. Soc. Pacific, Baltimore, 1997).
5. Angel, J. R. P. & Woolf, N. J. in *Planets Beyond the Solar System and the Next Generation of Space Missions* (ed. Soderblom, D.) 206–222 (Baltimore, 1997).
6. Angel, J. R. P. in *The Next Generation Space Telescope* (eds Bely, P. Y., Burrows, C. J. & Illingworth, G. D.) 81–94 (Space Telescope Science Institute, Baltimore, 1990).
7. Angel, J. R. P., Burge, J. H. & Woolf, N. J. Detection and spectroscopy of exo-planets like Earth. *Proc. SPIE* **2871**, 516–519 (1996).
8. Danchi, W. C., Bester, M., Degiacomi, C. G., Greenhill, L. J. & Townes, C. H. Characteristics of dust shells around 13 late-type stars. *Astron. J.* **107**, 1469–1513 (1994).
9. Heiden, J. C. *et al.* Two-dimensional images of alpha Orionis. *Astrophys. J.* **309**, 745–754 (1987).
10. Gilliland, R. L. & Dupree, A. K. First images of the surface of a star with the Hubble Space Telescope. *Astrophys. J.* **463**, L29–L32 (1996).
11. Lim, J., Carilli, C. L., White, S. M., Beasley, A. J. & Marson, R. G. Large convection cells as the source of Betelgeuse's extended atmosphere. *Nature* **392**, 575–577 (1998).
12. Bester, M. *et al.* Measurement at 11 micron wavelengths of the diameter of α Orionis and α Scorpii, and changes in effective temperature of α Orionis and very recent dust emission. *Astrophys. J.* **463**, 336–343 (1996).
13. Hinz, P. M. *et al.* First results of nulling interferometry with the Multiple Mirror Telescope. *Proc. SPIE* (in the press).
14. Noll, R. J. Zernike polynomial and atmospheric turbulence. *J. Opt. Soc. Am.* **66**, 207–211 (1976).
15. Hill, J. M. & Salinari, P. Large Binocular Telescope project. *Proc. SPIE* (in the press).
16. Angel, J. R. P. & Woolf, N. J. An imaging nulling interferometer to study extra-solar planets. *Astrophys. J.* **475**, 373–379 (1996).
17. Hoffmann, W. F., Hora, J. L., Fazio, G. G., Deutsch, L. K. & Dayal, A. MIRAC2, a mid-infrared camera for astronomy. *Proc. SPIE* (in the press).

Acknowledgements. We thank C. Foltz and the MMT staff for their essential support of the project. This work was supported by AFOSR, NASA and the US NSF.

Correspondence and requests for materials should be addressed to P.H. (e-mail: phinz@as.arizona.edu).

Superconductivity in a two-dimensional electron gas

Philip Phillips, Yi Wan, Ivar Martin, Sergey Knyshev & Denis Dalidovich

Loomis Laboratory of Physics, University of Illinois at Urbana-Champaign, 1100 W. Green Street, Urbana, Illinois 61801-3080, USA

In a series of recent experiments, Kravchenko and colleagues^{1,2} observed unexpectedly that a two-dimensional electron gas in zero magnetic field can become conducting at low temperatures: the two-dimensionality was imposed by confining the electron gas to the interface between two semiconductors. The observation of

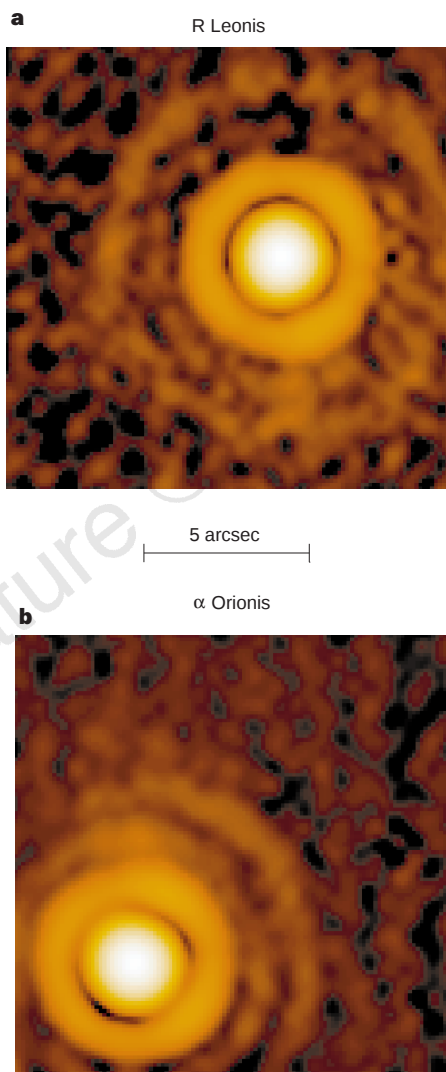


Figure 4 Stellar diffraction patterns obtained by subtracting the nulled images of Fig. 3 from the corresponding constructive images. For α Ori, a misalignment of the centroids by 0.3 arcsec was applied to obtain the symmetric diffraction pattern shown. The non-interfered images of unresolved stars, when similarly aligned and stacked, are not so perfect because they include the halo components from nebular emission or scattering due to higher-order wavefront aberrations. These subtract out in the difference image.

this conducting phase is surprising, as the conventional theory of metals precludes the existence of a metallic state at zero temperature in two dimensions³. Nevertheless, there are now several experiments confirming the existence of the new conducting phase in dilute two-dimensional electron gases in zero magnetic field^{4–7}. Here we argue, on the basis of an analysis of these experiments and general theoretical grounds, that this phase is in fact a superconductor with an inhomogeneous charge density.

Although the specific details differ as the semiconductor that confines the two-dimensional (2D) electron gas is changed, several important similarities^{1,2,4–7} exist among the reported observations of the transition to the conducting state: (1) the existence of a critical electron or hole density (n_c) above which the conducting phase appears, (2) a characteristic temperature, typically of the order of half the Fermi temperature, T_F , below which the resistivity on the conducting side decreases, (3) critical scaling, indicative of a quantum critical point, in the vicinity of the insulator/conducting-phase transition, (4) nonlinear current–voltage (I – V) curves that show a symmetry around the I – V curve at criticality, and (5) suppression of the conducting phase by a magnetic field. In samples¹ of silicon in metal oxide semiconductor field-effect transistors (MOSFETs), temperature (T) and electric field (E) scaling have made it possible to extract the dynamical and correlation length exponents, z and ν , respectively. By fitting the resistivity (ρ) measurements to functions of the form, $\rho(T, n) = f_1(|\delta|/T^b)$ and $\rho(E, n) = f_2(|\delta|/E^a)$, with $b = 1/z\nu$, $a = 1/(z+1)\nu$, $\delta = (n - n_c)/n_c$, and n the electron density, Kravchenko and colleagues¹ found that $\nu = 1.5$ and $z = 0.8$. Analogous scaling occurs in GaAs (ref. 5) but only in the most disordered samples. This indicates that the underlying transition in the clean system is first order, whereas in the presence of disorder it becomes second order and critical scaling applies.

Three postulates underlie the conventional theory of metals³: (1) Fermi liquid theory accurately describes the low-temperature physics of conventional clean metals, (2) the classical and quantum corrections to the conductivity are additive⁸ as the system size increases, and (3) the logarithmic derivative (β) of the dimensionless conductance with respect to the system size is a continuous monotonic single-valued function as the strength of the disorder is increased from weak to strong. From (2) and (3) it follows^{3,8} that β is always negative in two dimensions. Hence, as the system size increases, the conductance decreases and insulating behaviour necessarily occurs in two dimensions. The general applicability of the scaling analysis rests firmly on the extensive numerical and experimental work that has confirmed its central predictions and assumptions⁹.

Strictly speaking, however, Fermi liquid behaviour occurs only for $T \ll T_F$. Hence, the onset of a conducting phase at $T \approx T_F$ does not necessarily pose a paradox, provided that the resistivity turns upwards at sufficiently low temperatures. However, such behaviour is inconsistent with the experimental observations for two main reasons. First, there is no experimental indication that the resistivity turns upwards at low temperatures. Second, the critical scaling observed¹ signifies that the conductor–insulator transition is a genuine quantum phase transition.

The conventional theory of metals forces us then to approach the new conducting state in 2D by abandoning one or all of the postulates of the standard view. In the context of the experiments^{1–2,4–7}, the Coulomb interaction V_{ee} exceeds the Fermi energy, ϵ_F , by at least an order of magnitude. As Fermi liquid theory has only been proven to be valid for $V_{ee} \ll \epsilon_F$, it is the Fermi liquid assumption that is the most likely to be invalid in the limit of $V_{ee} \gg \epsilon_F$. It is for this reason that the experiments force us to consider non-Fermi liquid models. Chakravarty and co-workers¹⁰ proposed a Luttinger liquid state to explain the experimental observations. However, such a state has yet to be proven to exist except in one dimension. In fact in 2D, a $T = 0$ superconductor is

the only conducting non-Fermi liquid state proven to exist in the presence of disorder¹¹ and zero magnetic field. Hence, rather than speculate as to the existence of some yet-unproven non-Fermi liquid state in 2D, we take the conservative approach, and propose that superconductivity is the origin of the new conducting state in 2D.

We support this proposal with three subsidiary arguments. First, the generic features¹² of the conducting transition resemble those of known¹³ insulator–superconductor phase transitions. Second, magnetoresistance measurements^{1,5} provide evidence for the existence of a critical magnetic field above which the conducting phase is destroyed. Third, the conducting transition lies in close proximity to an incipient electron crystal state in which strong charge retardation effects could in principle lead to Cooper pair formation. Further, we draw a parallel between the 2D electron gas experiments and those experiments on inhomogeneous superconductors in which an apparent saturation of the resistivity has been observed at low temperatures^{14,15}. We argue that the enhanced role of classical fluctuations of the superconducting phase¹⁶ in 2D can significantly suppress the temperature at which the resistivity vanishes.

In support of the superconducting model are numerous transport measurements on the electron and hole systems. The value of the dynamical exponent¹³, $z = 1$, the nonlinearity in the I – V curves, as well as the observed temperature and electric field scaling of the resistivity are all consistent with what is observed in known insulator–superconductor transitions. We now analyse the experimental data on the magnetoresistance, and show that they offer evidence for a critical parallel magnetic field, consistent with singlet pairing. Although we limit our discussion to a parallel magnetic field ($H_{\parallel}$), Kravchenko and colleagues⁵ showed that the response of the conducting phase to magnetic field is independent of the direction of the field, indicating that it is a spin effect that destroys the conducting phase. Consider first the magnetoresistance measurements⁵ in GaAs. Measurements of the conductivity as a function of the hole density in the temperature interval 1.4–0.3 K reveal that even in the presence of a magnetic field, the conductivity curves cross at a unique value of the hole density. The single crossing point signifies that the conductivity is temperature-independent at a particular density. Hence, this density marks the transition between the conducting and insulating phases. If the conducting phase were destroyed by an arbitrarily small magnetic field, such a crossing point would not occur at finite field. However, the experiments show that even for fields as high as 1 T, a unique crossing point exists. By $H_{\parallel} = 3.0$ T, the unique crossing point vanishes, indicating that there is a threshold parallel magnetic field above which the conducting phase is extinguished. The critical field is, however, density dependent. From the crossing point in finite field, we conclude that for $\delta = 0.02$ and 0.03 , the critical fields are $H_{\parallel}^c = 0.5$ T and $H_{\parallel}^c = 1.0$ T, respectively. For GaAs, the Zeeman energies at such field strengths correspond to an energy at least an order of magnitude smaller than the Fermi temperature. Hence, the conducting phase in GaAs is characterized by an internal energy scale distinct from ϵ_F .

What about silicon MOSFETs? Based on the measurements of $\rho(H, E)$, Kravchenko *et al.*² concluded that the conducting phase is destroyed for an arbitrarily small parallel magnetic field, indicating a vanishing of the critical field. This conclusion poses a distinct problem if the conducting transition in Si MOSFETs and GaAs is assumed to be driven by the same physical process: that is, either both or neither should display a critical field (H_c). We point out, however, that the observation of a critical field in GaAs was based on an analysis of $\rho(H, T)$, not $\rho(H, E)$. If our interpretation is correct, $\rho(H, T)$ measurements on Si MOSFETs should also confirm the existence of a critical field. Hence, we investigate precisely what is contained in $\rho(H, T)$ for Si MOSFETs. The raw data indicates that for $H_{\parallel} \approx 9$ kOe the slope of $\rho(T)$ changes sign². Hence, from the raw data, there is no indication—at least within the temperature regime studied—that an arbitrarily small magnetic field suppresses the

conducting phase. As a consequence, we analysed $\rho(H, T)$ for a Si MOSFET (D. Simonian, S. Kravchenko, M. Sarachik, personal communication) with a scaling function of the form, $f(|H_{\parallel} - H_c|/T^{1/\alpha})$. In field-tuned 2D insulator–superconductor transitions^{13,17}, the resistivity scales on either side of H_c as a universal function¹⁷ of $|H - H_c|/T^{1/2\nu_B}$, where ν_B and ν_B are the dynamical and correlation length exponents, respectively, in a magnetic field. If a critical field exists, then above and below H_c the experimental values for the resistivity should collapse onto two distinct branches. Such collapse is shown in Fig. 1 above and below a critical field of $H_c = 9.5$ kOe with $\alpha = 0.6 \pm 0.1$. Here again, the critical field corresponds to an energy scale that is an order of magnitude smaller than the Fermi energy. We note that experiments on the insulator–superconductor transition in bismuth films (N. Markovic, C. Christiansen and A. M. Goldman, personal communication) show that $\nu_B \nu_B = 0.7 \pm 0.2$, which is remarkably close to the value, $\alpha = 0.6 \pm 0.1$, obtained in the scaling plot in Fig. 1.

The resilience of the conducting phase to a parallel magnetic field signifies that the ground state is a singlet. Although spin glass and antiferromagnetic order are also consistent with a singlet ground state, such phases are insulating in 2D. If we entertain the possibility of other non-Fermi liquid states, it is unclear how such a state would differ from a superconducting one, as the experiments dictate that such a state must also have a singlet energy gap and conduct in the presence of disorder.

In terms of the dimensionless measure of the density, $r_s = 1/(\pi \rho a_0^*)^{1/2}$ (where a_0^* is radius and ρ is the density), the conducting transition for Si MOSFETs¹ and GaAs (refs 5, 6) occurs at $r_s \approx 10$ and $r_s \approx 18$, respectively. For such dilute systems, $r_s \gg 1$, it is not known definitively what type of microscopic order occurs in the ground state. However, as a Fermi liquid description is valid for dense systems, $r_s \lesssim 1$, perturbation theory^{18,19} will necessarily fail to describe the experimental observations^{4,20}. Monte Carlo simulations²¹ reveal that in a clean 2D electron gas, a Wigner crystal is the ground state for $r_s > 37$. (A Wigner crystal is an electron crystalline phase in which the electrons minimize their repulsive potential energy and form an ordered array. For clean 2D systems, the electrons arrange themselves in a triangular lattice.) The simulations also indicate that in the presence of disorder^{22,23} or random pinning centres, an incipient Wigner crystal phase exhibiting quasi-long-range order still forms. The melting density for such

a phase is shifted to higher values, typically $r_s \approx 10$ (though the amount and type of disorder might change this value) because disorder stabilizes the solid relative to the liquid phase. As disorder is undoubtedly present in the experimental systems and the conducting transition occurs for $r_s \approx 10$, it makes sense to think about the transition to the conducting state as a transition from an insulating incipient Wigner crystal phase. Experimentally, the density dependence of the threshold electric field²⁴ required to initiate transport on the insulating side, $E_t \propto \delta^{1.5}$, is consistent with the prediction²⁵ for quantum tunnelling in an incipient Wigner crystal. Hence, experimental evidence also corroborates the proximity of the conducting transition to an insulating phase with quasi-long-range order.

The proximity of the conducting phase to a quasi-crystalline dilute electron phase makes this problem quite analogous to high-temperature superconductivity. Although this phenomenon is at present unexplained, experimentally it is clear that superconductivity in the copper oxides occurs as doped holes disrupt the perfect antiferromagnetic order in the Mott insulating phase. The correspondence with high T_c is even more striking as such systems are quasi-2D with $r_s \approx 10$ for the holes. In both systems, destruction of the long-range (or quasi-long-range) spin or charge order in the insulating phase is accompanied by a charge or spin retardation effect which attempts to preserve the memory of the correlations in the insulating phase. As an incipient Wigner crystal has at best frustrated antiferromagnetic²¹ spin order, the charge retardation effect is expected to dominate²⁶. We illustrate its role as follows.

Consider an electron crystalline, or quasi-crystalline, phase in which the electrons are locked into ‘home’ positions by the electron interaction. Such a crystalline phase is stable when the zero-point energy ω_0 much exceeds the kinetic energy, ϵ_F , of an electron in each unit cell (or Wigner–Seitz cell in the context of a Wigner crystal) of edge length r_s . A distinct feature of an electron crystal, or quasi-crystalline, phase is the dominant role played by the correlations. The correlations manifest themselves in the form of a correlation hole which is centred at the average position of each electron in a unit cell. Increasing the electron density leads to an increase in the electron kinetic energy and an eventual melting of a Wigner crystal. In the melted phase, correlation holes still form around electrons. However, because they are massive, relative to a single electron, their response is delayed. The retardation timescale is set by the inverse of

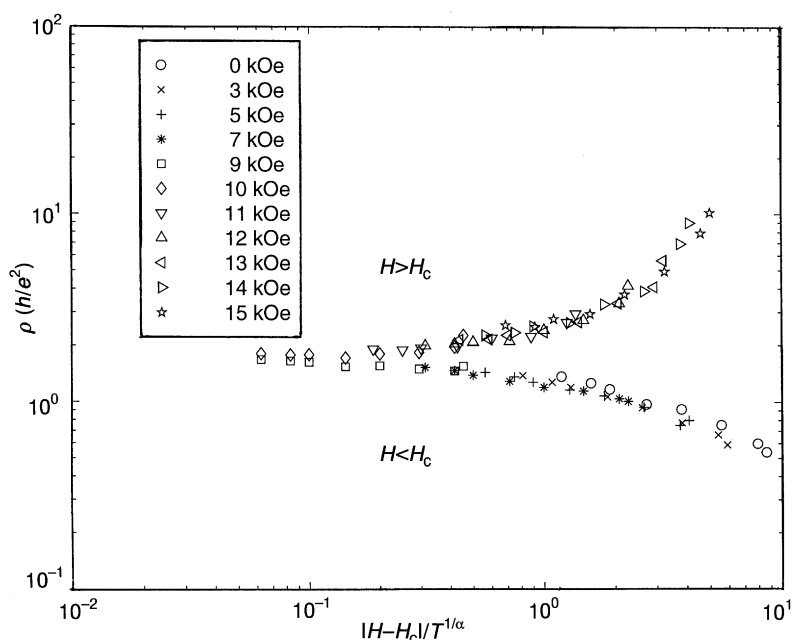


Figure 1 Scaling curve for the magnetoresistance obtained from experimental data of ref. 2. A critical field at 9.5 kOe with $\alpha = 0.6$ may be seen. Key shows magnetic field strength, H , corresponding to data points.

the plasma frequency. Within this timescale, a correlation hole lags behind its associated electron and hence appears positively charged to another electron. As a consequence, the partially vacated correlation hole can attract another electron. In so doing, the correlation hole mediates a dynamic attraction between electrons and the subsequent onset of Cooper pair formation. Because plasmons are ungapped in 2D, the retardation effect is strongest for the low-frequency plasmons. We anticipate that it is from such plasmons that the dominant electron attraction arises.

For a 3D electron gas, Takada²⁶ has made the analogous observation regarding the proximity of superconductivity to the melting of a Wigner crystal. The work of Kelly and Hanke²⁷ and Ren and Zhang²⁸ also relate to this proximity. More recently, Belitz and Kirkpatrick²⁹ have proposed a similar charge polarization mechanism in which the long-time tail of the charge density correlation function, which appears in the presence of disorder, assists pairing. Also, in the Monte Carlo simulations²¹ on a clean Wigner crystal, a precipitous drop of the spin susceptibility is observed in the melted phase, a prerequisite for singlet superconductivity. In addition, capacitance charging experiments on GaAs quantum dots³⁰, in which a 2D electron gas is constructed one electron at a time from the very first electron, have reported the occurrence of pair tunnelling events correspond to two electrons charging the same quantum state on the dot, such states form only if an electronic attraction screens the Coulomb repulsion between the two electrons. In so far as the electrons in a quantum dot realistically model a dilute ($r_s > 3$) 2D electron gas, it is certainly reasonable to suspect that the electron attraction also persists in Si MOSFETs.

In 2D superconductors, the resistivity vanishes not at the mean-field temperature, T_c , at which the pair amplitude is established, but at a lower temperature, T_0 , where global phase coherence obtains. That is, rather than defining the superconducting transition temperature, T_c determines the characteristic temperature at which the resistivity initially decreases. Pair fluctuations above T_0 are quenched by a magnetic field when the Zeeman and pairing energies are comparable. Both disorder³¹ and low superfluid densities¹⁶ can significantly lower T_0 relative to T_c in 2D systems. In fact, Emery and Kivelson¹⁶ have argued that a crucial feature which suppresses T_0 in high- T_c materials is their notoriously low superfluid density. In the electron systems of interest, disorder is present (leading to inhomogeneous electron density) and the superfluid density is expected to be small as the electron gas is dilute. As a consequence, we anticipate that in Si MOSFETs and GaAs, T_0 will be greatly suppressed relative to the mean-field T_c . In the case of GaAs (refs 5, 6), an exponential drop of the resistivity on the conducting side is accompanied by a plateau below some characteristic temperature. This latter experimental trend is identical to the behaviour observed by Goldman and collaborators¹⁴ and Imry and Strongin¹⁵ in inhomogeneous 2D superconductors. The strong similarity between the inhomogeneous systems and the GaAs data suggests that the phase-locked zero-resistance state occurs at a temperature significantly lower than that probed experimentally. Nonetheless, the nonlinear I - V characteristics¹ observed in the Si MOSFETs is

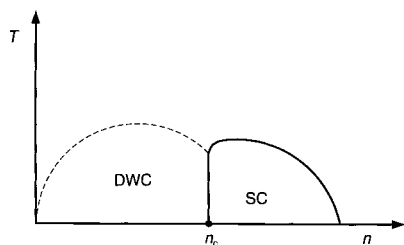


Figure 2 Schematic mean-field phase diagram for a 2D disordered electron system. DWC, disordered Wigner crystal; SC, superconductor. The filled circle at n_c is determined from the experimental data in ref. 1.

consistent with a 2D superconducting state on the brink of global phase coherence.

In the dense limit, correlation and retardation effects are negligible, and consequently Cooper pair formation ceases. The anticipated termination of the electron attraction at small r_s suggests the schematic phase diagram depicted in Fig. 2. Beyond the upper density at which superconducting pair fluctuations cease, the electron liquid is probably insulating. Hence, a direct transition from an insulator to a metal in 2D by changing the density seems unlikely. In GaAs samples, a re-entrant (M. Y. Simmons, unpublished data) insulating phase at $r_s \approx 8$ was observed, consistent with the analysis here. Further experiments are needed on the MOSFETs to see the re-entrant insulating phase.

Because Fermi liquids are localized in 2D, the new conducting state must be some type of non-Fermi liquid. As $T = 0$ superconductivity is the only conducting, disordered, non-Fermi liquid proven to exist in 2D, this must be viewed as the leading candidate to explain the experimental observations. The existence of a critical parallel magnetic field is a clear indicator of a singlet energy gap in the conducting phase in 2D. However, further experiments are needed to probe the low-temperature physics as well as the magnetoresistance in Si MOSFETs. In addition to disorder and a suppressed superfluid density, the inhomogeneities introduced in a dilute electron gas ($r_s > 3$) as a result of the negative compressibility³² (an instability relative to a uniform charge density) are expected to enhance phase fluctuations as well. Consequently, we anticipate that the transition to the phase-locked state will occur at a temperature significantly below the mean-field T_c . □

Received 19 May; accepted 8 July 1998.

1. Kravchenko, S. V., Simonian, D., Sarachik, M. P., Mason, W. & Furneaux, J. E. Electric field scaling at a $B=0$ metal-insulator transition in two dimensions. *Phys. Rev. Lett.* **77**, 4938–4941 (1996).
2. Simonian, D., Kravchenko, S. V. & Sarachik, M. P. Magnetic field suppression of the conducting phase in two dimensions. *Phys. Rev. Lett.* **79**, 2304–2307 (1997).
3. Abrahams, E., Anderson, P. W., Licciardello, D. C. & Ramakrishnan, T. V. Scaling theory of localization: absence of quantum diffusion in two dimensions. *Phys. Rev. Lett.* **42**, 673–676 (1979).
4. Popovic, D., Fowler, A. B. & Washburn, S. Metal-insulator transition in two dimensions: effects of disorder and magnetic field. *Phys. Rev. Lett.* **79**, 1543–1546 (1997).
5. Simmons, M. Y. *et al.* Metal-insulator transition at $B=0$ in a dilute two dimensional GaAs/AlGaAs hole gas. *Phys. Rev. Lett.* **80**, 1292–1295 (1998).
6. Hanien, Y. *et al.* Metallic-like conductivity of a two-dimensional hole system. *Phys. Rev. Lett.* **80**, 1288–1291 (1998).
7. Coleridge, P. T. *et al.* Metal-insulator transition at $B=0$ in p-SiGe. Preprint cond-mat/9708118 at (<http://xxx.lanl.gov>) (1997).
8. Langer, J. S. & Neal, T. Breakdown of the concentration expansion for the impurity resistivity of metals. *Phys. Rev. Lett.* **16**, 984–986 (1966).
9. Lee, P. A. & Ramakrishnan, A. R. Disordered electronic systems. *Rev. Mod. Phys.* **57**, 287–337 (1985).
10. Chakravarty, S., Yin, L. & Abrahams, E. Interactions and scaling in a disordered 2D metal. Preprint cond-mat/9712217 at (<http://xxx.lanl.gov>) (1997).
11. Anderson, P. W. Theory of dirty superconductors. *J. Phys. Chem. Solids* **11**, 26–30 (1959).
12. Phillips, P. & Wan, Y. Incipient p-wave superconductivity in a Si MOSFET. Preprint cond-mat/9704200 at (<http://xxx.lanl.gov>) (1997).
13. Yazdani, A. & Kapitulin, A. Superconducting-insulating transition in two-dimensional a-MoGe thin films. *Phys. Rev. Lett.* **74**, 3037–3040 (1995).
14. Jaeger, H. M. *et al.* Onset of superconductivity in ultrathin granular metal films. *Phys. Rev. B* **40**, 182–196 (1989).
15. Imry, Y. & Strongin, M. Destruction of superconductivity in granular and highly disordered metals. *Phys. Rev. B* **24**, 6353–6360 (1981).
16. Emery, V. J. & Kivelson, S. A. Importance of phase fluctuations in superconductors with small superfluid density. *Nature* **374**, 434–437 (1995).
17. Fisher, M. P. A. Quantum phase transitions in disordered two-dimensional superconductors. *Phys. Rev. Lett.* **65**, 923–926 (1990).
18. Castellani, C., DiCastro, C. & Lee, P. A. Metallic phase and metal-insulator transition in 2d electronic systems. *Phys. Rev. B* **57**, R9381–R9384 (1998).
19. Dobrosavljevic, V., Abrahams, E., Miranda, E. & Chakravarty, S. Scaling theory of two-dimensional metal-insulator transitions. *Phys. Rev. Lett.* **79**, 455–458 (1997).
20. Pudalov, V. M., Brunthaler, G., Prinz, A. & Bauer, G. Logarithmic temperature dependence of the conductivity and lack of universal one-parameter scaling in the two-dimensional metal. Preprint cond-mat/9801077 at (<http://xxx.lanl.gov>) (1998).
21. Tanatar, B. & Ceperley, D. M. Ground state of the two-dimensional electron gas. *Phys. Rev. B* **39**, 5005–5016 (1989).
22. Eguluz, A. G., Maradudin, A. A. & Elliott, R. J. Two-dimensional Wigner lattice in a magnetic field and in the presence of a random array of pinning centers. *Phys. Rev. B* **27**, 4933–4945 (1983).
23. Chui, S. T. & Tanatar, B. Impurity effect on the two-dimensional-electron fluid-solid transition in zero field. *Phys. Rev. Lett.* **74**, 458–461 (1995).
24. Pudalov, V. M., D'Iorio, M., Kravchenko, S. V. & Campbell, J. W. Zero-magnetic-field collective insulator phase in a dilute 2D electron system. *Phys. Rev. Lett.* **70**, 1866–1869 (1993).
25. Chui, S. T. Depinning, defect creation and quantum tunneling. *Phys. Lett. A* **180**, 149–153 (1993).
26. Takada, Y. s- and p-wave pairings in a dilute electron gas. *Phys. Rev. B* **47**, 5202–5211 (1993).
27. Kelly, M. J. & Hanke, W. Surface superconductivity and the MOS system. *Phys. Rev. B* **23**, 112–123 (1981).

28. Ren, Y. & Zhang, F. C. Fermion analogy of anyon superconductivity in the two-dimensional electron gas. *Phys. Rev. B* **49**, 1532–1535 (1994).
29. Belitz, D. & Kirkpatrick, T. Are Si MOSFET's triplet superconductors. Preprint cond-mat/9705023 at (<http://xxx.lanl.gov>) (1997).
30. Zhitenev, N. B., Ashoori, R. C., Pfeiffer, L. N. & West, K. W. Periodic and aperiodic bunching in the addition spectra of quantum dots. *Phys. Rev. Lett.* **77**, 1833–1836 (1996).
31. Beasley, M. R., Mooij, J. E. & Orlando, T. P. Possibility of vortex-antivortex pair dissociation in two-dimensional superconductors. *Phys. Rev. Lett.* **42**, 1165–1168 (1979).
32. Moroni, S., Ceperley, D. & Senatore, G. in *Proc. Rochester Symp. on Strongly Coupled Plasmas* (eds van Horn, H. & Ichimaru, S.) 445–452 (Univ. Rochester Press, Rochester, 1993).

Acknowledgements. We thank B. Laughlin for comments on the presentation of this Letter, and thank E. Fradkin, N. Markovic, A. Goldman, A. Yazdani, A. Castro-Neto, A. Leggett, D. Ceperley, T. Giamarchi, P. Parris and S. Wan for discussions. This work was supported by the ACS Petroleum Research Fund and the NSF.

Correspondence and requests for materials should be addressed to P.P. (e-mail: dimer@uiuc.edu).

Laminated fabrication of polymeric photovoltaic diodes

M. Granström*, K. Petritsch*, A. C. Arias*, A. Lux†, M. R. Andersson‡ & R. H. Friend*

* Cavendish Laboratory, Department of Physics, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

† Cambridge Display Technology Ltd, 181A Huntingdon Road, Cambridge CB3 0DJ, UK

‡ Department of Polymer Technology, Chalmers University of Technology, S-412 96 Gothenburg, Sweden

Photoexcited electron transfer between donor and acceptor molecular semiconductors provides a method of efficient charge generation following photoabsorption, which can be exploited in photovoltaic diodes^{1–3}. But efficient charge separation and transport to collection electrodes is problematic, because the absorbed photons must be close to the donor–acceptor heterojunction, while at the same time good connectivity of the donor and acceptor materials to their respective electrodes is required. Mixtures of acceptor and donor semiconducting polymers^{3,4} (or macromolecules⁵) can provide phase-separated structures which go some way to meeting this requirement, providing high photoconductive efficiencies. Here we describe two-layer polymer diodes, fabricated by a lamination technique followed by controlled annealing. The resulting structures provide good connectivity to the collection electrodes, and we achieve a short-circuit photovoltaic quantum efficiency of up to 29% at optimum wavelength, and an overall power conversion efficiency of 1.9% under a simulated solar spectrum. Given the convenience of polymer processing, these results indicate a promising avenue towards practical applications for such devices.

Applications involving conjugated polymers include light-emitting diodes, thin-film transistors, sensors and photovoltaic devices^{6–10}. The last use the photovoltaic effect, by which electrons and holes are produced, and collected at electrodes when a semiconductor device is illuminated. In inorganic semiconductors, the photogenerated electrons and holes are free, whereas in the organic materials the absorption creates bound electron–hole pairs (excitons). These can be dissociated by the use of combinations of materials, one with small ionization potential and good hole-transport properties and the other with large electron affinity that acts as electron-transport material. This route has proven successful for organic photovoltaic devices, both as multilayer devices^{11–13} and single-layer blend devices. The blends can be either polymer blends^{3,4} or blends created by mixing molecular substances, such as C₆₀ into a polymer matrix^{5,14}. If this can be done so that a percolation path for the charge carrier is created, the efficiency of the devices is increased dramatically¹⁵.

We use here a cyano derivative of poly(*p*-phenylene vinylene), MEH-CN-PPV (refs 16–18), as electron acceptor, and a derivative

of polythiophene, POPT (refs 19–21), as hole acceptor. The structures of these are shown in Fig. 1a. The two polymers are soluble in chloroform or toluene, and are therefore conveniently mixed in solution. The structure of the laminated diodes is shown in Fig. 1b; these are assembled by forming a POPT-rich film on the ITO- or PEDOT-coated glass and an MEH-CN-PPV-rich film on the aluminium or calcium-coated glass substrate, and performing the lamination at an elevated temperature (here ITO is indium tin oxide; PEDOT is described in Fig. 1 legend).

If exciton splitting occurs in a blend of fluorescent polymers, it is usually faster than its radiative decay and can therefore be observed as photoluminescence (PL) quenching³. We measured the absolute PL efficiency using an integrating sphere as described previously²², using excitation of wavelength 488 nm. In the MEH-CN-PPV/POPT blend, the PL is strongly quenched with only 5% or 10% of the thiophene polymer in the blend, dropping from 44% absolute PL efficiency for a pure MEH-CN-PPV film down to as low as 2%; this indicates efficient exciton dissociation in the blend. Energy

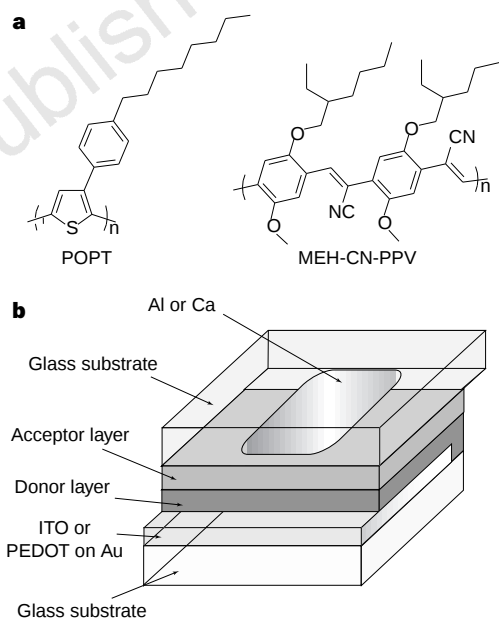


Figure 1 Chemical structures and device structure. **a**, Chemical structures of the polymers used in the devices. POPT is a regioregular phenyl-octyl substituted polythiophene which is found in two different forms^{20,21} with the red-shifted structure obtained by thermal annealing of spin-coated films. This red-shifted form is preferred here as it provides the best match to the solar spectrum. MEH-CN-PPV is a fluorescent cyano-derivative of poly(*p*-phenylene vinylene), with a large electron affinity as a result of the electron-withdrawing cyano groups¹⁶. It and related polymers can be used as the emissive materials in efficient polymer light-emitting diodes^{17,18}. **b**, Device structure. Both polymers (see **a** for chemical structures) were dissolved in chloroform or toluene (5 mg ml⁻¹) and the solutions were filtered with 0.45-μm filters before use. For the top half of the device, aluminium or calcium contacts were thermally evaporated ($P < 10^{-6}$ torr) on glass substrates, and the acceptor material, MEH-CN-PPV (and a small amount of POPT, usually 5%), was spin-coated on top of the metal electrode. The other half of the device comprises the donor material, POPT (and a small amount of MEH-CN-PPV, usually 5%), spin-coated on either ITO substrates or glass coated with poly(ethylene dioxythiophene) doped with polystyrene sulphonic acid (PEDOT:PSS, Bayer AG, Ludwigshafen, Germany). To ensure a low contact resistance, a thin layer of gold (~10 nm) was thermally evaporated on the glass slide before spin-coating the PEDOT film from a water solution. The thickness of the PEDOT layer was 100 nm. To get the desired structure of the POPT¹⁹, this half of the device was heated to 200°C under vacuum before the device was laminated together by applying a light pressure while one half was still at elevated temperature. The total thickness of the organic semiconductive layer was 70–80 nm, and the active area 2.5 mm².

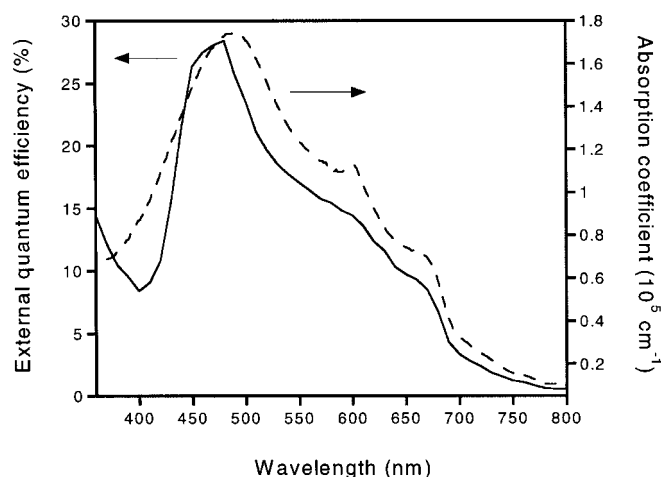


Figure 2 Optical properties. Shown are the optical absorption, measured as $-\log(\text{transmittance})$, of a POPT/MEH-CN-PPV double layer (dashed line) and external quantum efficiency of an ITO/POPT:MEH-CN-PPV (19:1)/MEH-CN-PPV:POPT (19:1)/Al device (solid line). Measurements of photocurrent were made by illuminating the samples from the ITO or PEDOT side with a tungsten-halogen lamp, dispersed by a Bentham M300 single-grating monochromator. Electrical data were measured and bias voltage was applied using a Keithley 237 source-measure unit. Absorption spectra were measured with a Hewlett Packard 8453 ultraviolet-visible spectrometer.

transfer to the blending partner as an alternative explanation is ruled out because the PL of both materials is quenched in the blend.

As a starting point, we made single-layer blend devices, similar to those previously reported^{3,4}, from the MEH-CN-PPV/POPT blends. This was done with non-heated films, resulting in devices with external quantum efficiencies of the order of 3–5%, which is comparable to other polymer blend devices. However, because POPT needs to be heated (or solvent-treated) to reach the ordered state which has its absorption in the long-wavelength range, the phase-separation properties become less favourable. Because of the heating, not only do the POPT chains themselves have a chance to relax and find a more ordered state, so does the blend as a whole. This results in a phase separation on a much larger length scale than preferred (about 200–400 nm, as seen in atomic force microscopy (AFM) images), which radically decreases the external quantum efficiency down to about 0.05%, a value comparable to the result for single-layer MEH-CN-PPV devices.

We developed the device structures using laminated double-layer arrangements (Fig. 1b). In this way, it is possible to reach the desired coverage of the wavelength range, and, simultaneously, make sure that the acceptor material has a proper contact with the low-work-function contact (Al or Ca) and the donor with the high-work-function contact (ITO or PEDOT). Simple laminated devices with two homolayers were found to have the desired wavelength response, but not particularly high efficiencies, albeit somewhat higher than the single-layer blend devices. But, by adding a small amount of POPT to the MEH-CN-PPV layer, 2–5 wt%, the efficiency of the devices was much increased compared to the double-layer device or single-layer blend. The external quantum efficiency follows the absorption (Fig. 2), suggesting that there is no inactive layer which would act as an optical filter and thereby decrease the efficiency. At -1 V and -2 V the curve (not shown) is slightly shifted towards shorter wavelengths, and the efficiencies at 480 nm are 49% and 58%, respectively. The addition of a similar small amount of MEH-CN-PPV to the POPT layer did not make any significant difference to the device efficiency, most probably because of phase-separation changes caused by heating the film.

Figure 3 shows the current–voltage response of the device, measured in darkness and excited at 480 nm. A distinct diode

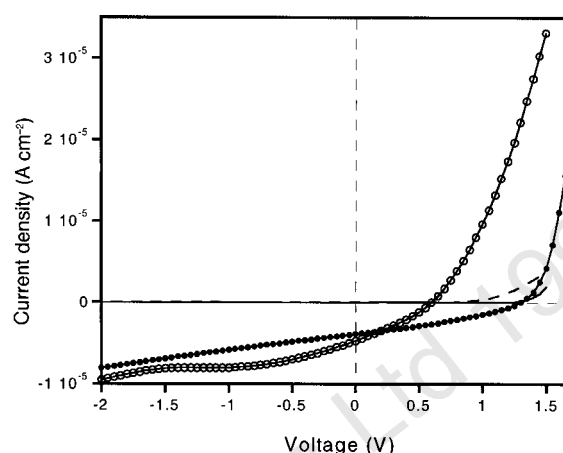


Figure 3 Current–voltage characteristics. Data shown are for laminated double-layer devices using POPT:MEH-CN-PPV (19:1)/MEH-CN-PPV:POPT (19:1) blends. Open circles and dashed line show light (480 nm, $34 \mu\text{W cm}^{-2}$) and dark current for a device with ITO and Al as electrodes (open-circuit voltage 0.6 V), whereas the filled circles and solid line show the response in light and darkness for a device with PEDOT and Ca as electrode materials (open-circuit voltage 1.3 V). The characteristics were measured in the dark, and using narrow lines from the monochromator.

behaviour is found, and the rectification ratio in the dark is $>10^3$ at ± 1.5 V. The fill factor, FF, is defined as the maximum power $(IV)_{\text{max}}$ that can be extracted from a photovoltaic diode divided by the product of maximum current I_{sc} and maximum voltage V_{oc} :

$$\text{FF} = \frac{(IV)_{\text{max}}}{I_{\text{sc}} V_{\text{oc}}} \quad (1)$$

Using this definition, we get values in the range of 30–35% for both types of devices described above (single-layer blend and laminated double-layer blend). The power conversion efficiency η_p is defined as the ratio of the input power P_{in} , in the form of electromagnetic radiation, to the output power P_{out} in the form of electrical energy:

$$\eta_p = \frac{P_{\text{out}}}{P_{\text{in}}} = \frac{(IV)_{\text{max}}}{LA} \quad (2)$$

where L is the intensity of the illumination, and A is the active area. When the fill factor is known, equation (2) translates into:

$$\eta_p = \text{FF} \frac{I_{\text{sc}} V_{\text{oc}}}{LA} \quad (3)$$

Open-circuit voltage can be selected by choice of the work functions of the electron- and hole-accepting electrodes. We have maximized this for an Au/PEDOT/POPT:MEH-CN-PPV (19:1)/MEH-CN-PPV:POPT (19:1)/Ca device, for which the current–voltage characteristics are shown, together with those for a device made with ITO and Al electrodes, in Fig. 3. The spectral response is very similar for both devices. For the Au/PEDOT and Ca electrode device, the external quantum efficiency reaches 29% with 480 nm excitation and no applied bias. Calculating the power conversion efficiency using these values then gives $\eta_p = 4.8\%$ at the peak wavelength, rising to $\sim 7\%$ at solar illumination intensities, which is a significant increase compared to previous polymer devices. We calculate a power conversion efficiency for a simulated solar spectrum, AM 1.5 (77 mW cm^{-2}) of 1.9%. Under 2 V negative bias, this device has a quantum efficiency of 61% at the peak wavelength. The dependence of open-circuit voltage and short-circuit current on the intensity of incident light L (at 488 nm) is shown in Fig. 4. The short-circuit current is very close to linear ($I_{\text{sc}} \propto L^{1.02}$) up to the highest value measured here of 100 mA cm^{-2} , whereas the open-circuit voltage

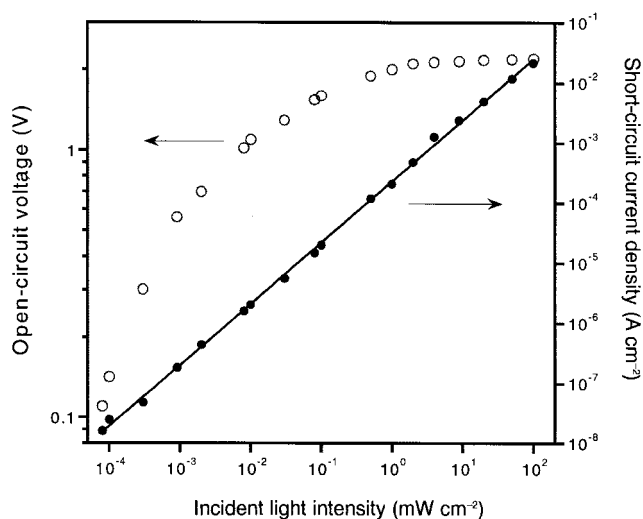


Figure 4 Effect of light intensity. Shown are the dependence of open-circuit voltage (open circles) and short-circuit current density (filled circles) of an Au/PEDOT/POPT:MEH-CN-PPV (19:1)/MEH-CN-PPV:POPT (19:1)/Ca device on incident light intensity. The device was illuminated using the 488-nm line from an argon-ion laser. The fill factor, as defined in equation (1), is relatively constant, ranging from 0.30 to 0.34.

increases sublinearly with intensity, saturating when the value approaches 2.2 V, which is close to the estimated difference in work function between the two electrodes.

We used tapping-mode AFM to investigate the structure of these devices. This technique has been found previously to give good contrast between different polymers²³, and we find here that better contrast is obtained using phase-detection (which provides a measure of the viscoelasticity). Figure 5a and b shows the in-plane structures of the MEH-CN-PPV-rich layer and the POPT-rich layers, respectively. We note clear evidence for formation of 'islands' of the minority phase, which are larger in size for the heat-treated POPT-rich film. Figure 5c shows the cross-section of a laminated structure: there is interpenetration between the two layers following the lamination and annealing procedure, with a length scale of 20–30 nm. We note that finer-scale interpenetration is not expected to be revealed in these AFM images.

There is increased interest in solar energy conversion, though cost per installed kilowatt for inorganic semiconductor devices remains high, and there is renewed interest in the development of novel photovoltaic cells. These include photoelectrochemical cells which use dye adsorbed onto sintered titanium dioxide, which show high solar energy conversion efficiencies²⁴ and which are under active development. The devices which we have developed now show quantum efficiencies which approach usable values; they also have several desirable practical attributes such as low-temperature processing which is compatible with fabrication onto polymer substrates, which offers low-cost large-area manufacture. Our devices have good efficiency, especially at high illumination levels (Fig. 4), which is achieved in spite of low carrier mobilities because the total semiconductor film thickness is very small (100 nm here). We also comment that the devices which we report here show efficient operation over the full visible spectrum, and, by virtue of the absorption spectrum of the POPT, even extends into the infrared region. Further development is required to match better the solar spectrum. □

Received 6 April; accepted 19 June 1998.

1. Tang, C. W. Two-layer organic photovoltaic cell. *Appl. Phys. Lett.* **48**, 183–185 (1986).
2. Halls, J. J. M., Pichler, K., Friend, R. H., Moratti, S. C. & Holmes, A. B. Exciton diffusion and dissociation in a poly(p-phenylenevinylene)/C₆₀ heterojunction photovoltaic cell. *Appl. Phys. Lett.* **68**, 3120–3122 (1996).

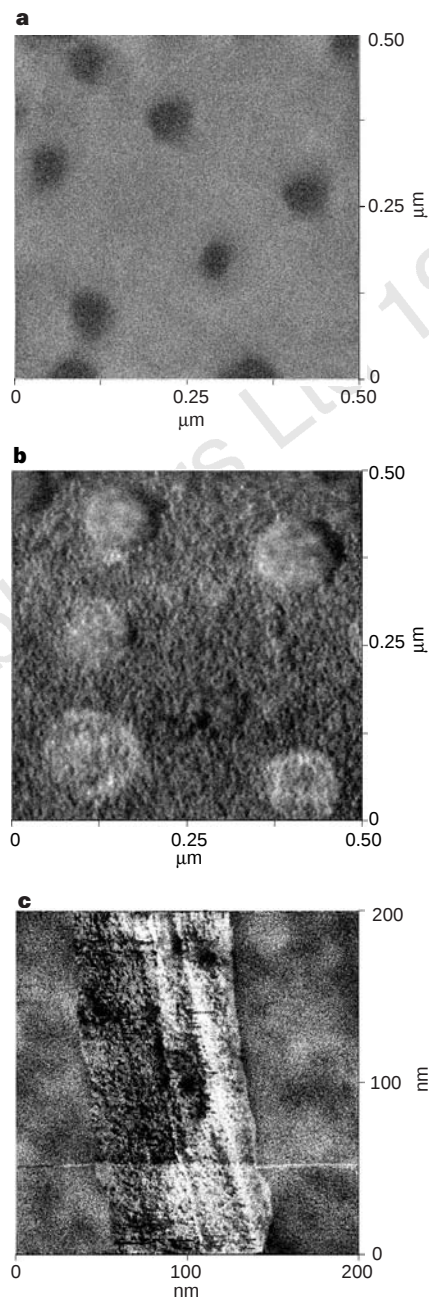


Figure 5 Tapping-mode AFM images. **a**, An in-plane image of MEH-CN-PPV:POPT (19:1). **b**, An in-plane image of annealed POPT:MEH-CN-PPV (19:1). The images are shown in phase contrast, most of which is due to differences in viscoelastic properties of the two polymers; height differences are no more than 5 nm. **c**, A cross-sectional image, obtained for a structure formed without metal electrodes and with polyester substrates (the cut to reveal the cross-section was made at low temperature). Images were obtained using phase-sensitive detection, and were taken using a NanoScope IIIa Dimension 3100 (Digital Instruments Inc, Santa Barbara).

3. Halls, J. J. M. *et al.* Efficient photodiodes from interpenetrating polymer networks. *Nature* **376**, 498–500 (1995).
4. Yu, G. & Heeger, A. J. Charge separation and photovoltaic conversion in polymer composites with internal donor-acceptor heterojunctions. *J. Appl. Phys.* **78**, 4510–4515 (1995).
5. Yu, G., Gao, J., Hummelen, J. C., Wudl, F. & Heeger, A. J. Polymer photovoltaic cells—enhanced efficiencies via a network of internal donor-acceptor heterojunctions. *Science* **270**, 1789–1791 (1995).
6. Burroughes, J. H. *et al.* Light-emitting diodes based on conjugated polymers. *Nature* **347**, 539–541 (1990).
7. Willander, M., Assadi, A. & Svensson, C. Polymer-based devices their function and characterization. *Synth. Met.* **57**, 4099–4104 (1993).
8. Partridge, A. C., Harris, P. & Andrews, M. K. High-sensitivity conducting polymer sensors. *Analyst* **121**, 1349–1353 (1996).

9. Yu, G. & Heeger, A. J. High efficiency photonic devices made with semiconducting polymers. *Synth. Met.* **85**, 1183–1186 (1987).
10. Friend, R. H. *et al.* Electronic excitations in luminescent conjugated polymers. *Solid State Commun.* **102**, 249–258 (1997).
11. Tada, K., Onoda, M., Zakhidov, A. A. & Yoshino, K. Characteristics of poly(p-pyridyl vinylene)/poly(3-alkylthiophene) heterojunction photocell. *Jpn. J. Appl. Phys. Pt2-Lett.* **36**, L306–L309 (1997).
12. Takahashi, K. *et al.* Enhanced quantum yield in porphyrin/electron-donor double-layer solar cells. *Solar Energy Mater. Solar Cells* **45**, 127–139 (1997).
13. Yoshino, K., Tada, K., Fujii, A., Conwell, E. M. & Zakhidov, A. A. Novel photovoltaic devices based on donor-acceptor molecular and conducting polymer systems. *IEEE Trans. Electron. Devices* **44**, 1315–1324 (1997).
14. Roman, L. S., Andersson, M. R., Yohannes, T. & Inganäs, O. Photodiode performance and nanostructure of poly(thiophene)/C₆₀ blends. *Adv. Mater.* **9**, 1164–1168 (1997).
15. Yang, C. Y. & Heeger, A. J. Morphology of composites of semiconducting polymers mixed with C₆₀. *Synth. Met.* **83**, 85–88 (1996).
16. Moratti, S. C. *et al.* High electron-affinity polymers for LEDs. *Synth. Met.* **71**, 2117–2120 (1995).
17. Greenham, N. C., Moratti, S. C., Bradley, D. D. C., Friend, R. H. & Holmes, A. B. Efficient polymer-based light-emitting diodes based on polymers with high electron affinities. *Nature* **365**, 628–630 (1993).
18. Baigent, D. R. *et al.* Light-emitting diodes fabricated with conjugated polymers—recent progress. *Synth. Met.* **67**, 3–10 (1994).
19. Andersson, M. R. *et al.* Electroluminescence from substituted poly(thiophenes)—from blue to near-infrared. *Macromolecules* **28**, 7525–7529 (1995).
20. Berggren, M. *et al.* Thermal control of near-infrared and visible electroluminescence in alkyl-phenyl substituted polythiophenes. *Appl. Phys. Lett.* **65**, 1489–1491 (1994).
21. Andersson, M. R. *et al.* Regioselective polymerization of 3-(4-octylphenyl)thiophene with FeCl₃. *Macromolecules* **27**, 6503–6506 (1994).
22. de Mello, J. C., Wittmann, H. F. & Friend, R. H. An improved experimental determination of external photoluminescence quantum efficiency. *Adv. Mater.* **9**, 230–232 (1997).
23. Granström, M. *et al.* Self-organizing polymer films—a route to novel electronic devices based on conjugated polymers. *Supramol. Sci.* **4**, 27–34 (1997).
24. O'Regan, B. & Grätzel, M. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO₂ films. *Nature* **353**, 737–740 (1991).

Acknowledgements. Financial support from the Engineering and Physical Sciences Research Council, the European Commission (TMR Marie Curie fellowship and TMR Network SELOA), and from CNPq, Brazilian government, is gratefully acknowledged.

Correspondence and requests for materials should be addressed to R.H.F. (e-mail: rhf10@cam.ac.uk).

RNA-catalysed nucleotide synthesis

Peter J. Unrau & David P. Bartol

Whitehead Institute for Biomedical Research, and Department of Biology, MIT, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA

The 'RNA world' hypothesis proposes that early life developed by making use of RNA molecules, rather than proteins, to catalyse the synthesis of important biological molecules¹. It is thought, however, that the nucleotides constituting RNA were scarce on early Earth^{1–4}. RNA-based life must therefore have acquired the ability to synthesize RNA nucleotides from simpler and more readily available precursors, such as sugars and bases. Plausible prebiotic synthesis routes have been proposed for sugars⁵, sugar phosphates⁶ and the four RNA bases^{7–11}, but the coupling of these molecules into nucleotides, specifically pyrimidine nucleotides, poses a challenge to the RNA world hypothesis^{1–3}. Here we report the application of *in vitro* selection to isolate RNA molecules that catalyse the synthesis of a pyrimidine nucleotide at their 3' terminus. The finding that RNA can catalyse this type of reaction, which is modelled after pyrimidine synthesis in contemporary metabolism, supports the idea of an RNA world that included nucleotide synthesis and other metabolic pathways mediated by ribozymes.

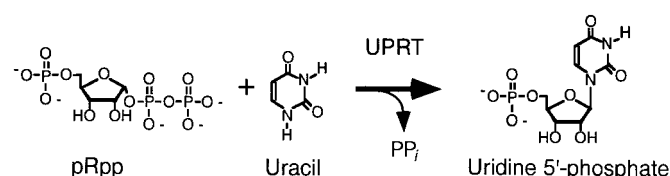


Figure 1 UPRT-catalysed synthesis of uridine 5'-phosphate from pRpp and uracil.

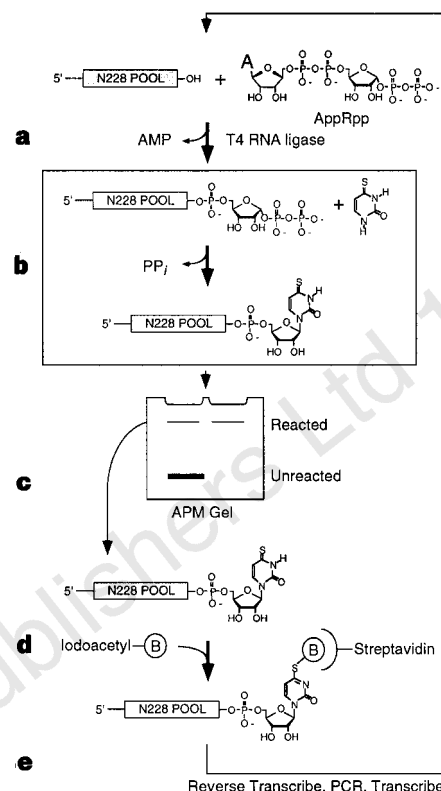


Figure 2 *In vitro* selection scheme. Symbols: PP_i, pyrophosphate; B, biotin. **a**, Tethering pool RNA to pRpp by using T4 RNA ligase and adenylated pRpp (AppRpp). **b**, ⁴⁵U synthesis promoted by active sequences within the initial pool. **c**, Enrichment of reacted sequences by using APM gels. One lane (left) contained radiolabelled pool RNA. The other lane (right) contained radiolabelled synthetic standard to mark the location of the reacted RNA. After the first round of selection, catalysts were enriched by serial purification on two APM gels. **d**, Further enrichment of ⁴⁵U-containing sequences by derivation with iodoacetyl-LC-biotin (Pierce) and capture with streptavidin magnetic beads. **e**, Amplification of enriched RNA by reverse transcription, PCR and transcription. The amplified RNA was then subjected to another round of selection *in vitro*.

In modern metabolism, pyrimidine nucleotides are synthesized from activated ribose (pRpp) and pyrimidine bases (such as uracil or orotate). For example, uracil phosphoribosyltransferase (UPRT) catalyses the reaction shown in Fig. 1. The chemistry of this reaction, nucleophilic attack on carbon after release of pyrophosphate, is central to the biosynthesis of nucleotides and amino acids (histidine and tryptophan), yet absent from known ribozyme reactions. The reaction differs from known RNA-catalysed reactions in other key respects: it occurs by an S_N1 mechanism (involving the stabilization of an oxocarbenium at the reaction centre, C1' of ribose)^{12,13}, and uracil is significantly smaller than the smallest ribozyme substrates.

To explore the ability of RNA to promote nucleotide synthesis, we performed an iterative selection *in vitro* to isolate from random sequences ribozymes that synthesized a pyrimidine nucleotide at their 3' terminus (Fig. 2). The initial pool of sequences contained > 1.5 × 10¹⁵ different RNA molecules, each with 228 random positions. To begin each selective round, pRpp was attached to the 3' end of pool RNA. RNA-pRpp conjugates were incubated with a uracil analogue, 4-thiouracil (⁴⁵Ura), to allow those sequences capable of glycosidic bond formation the opportunity to link their tethered ribose to ⁴⁵Ura. RNAs attached to the newly synthesized nucleotide, 4-thiouridine (⁴⁵U), were then enriched, amplified, and subjected to another round of selection–amplification. ⁴⁵U was chosen as the desired product because the thione at position 4

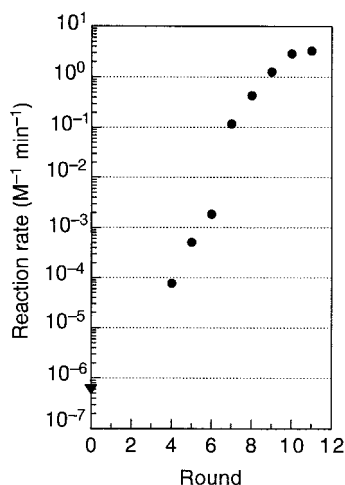


Figure 3 Increased ribozyme activity with successive rounds of selection. The upper bound for the uncatalysed rate is also plotted (triangle). Rounds 4 to 6 included error-prone PCR amplification, as described¹⁷. From rounds 7 to 10 the stringency of the selection was increased exponentially by decreasing both ⁴⁵Ura concentration (from 8 mM to 40 μ M) and incubation time (from 18 h to 7.5 min). Reaction rates were calculated by dividing the observed initial rate by the ⁴⁵Ura incubation concentration.

interacts strongly with a number of thiophilic reagents¹⁴, facilitating the efficient enrichment of RNAs with a single ⁴⁵U nucleotide. In other respects, sulphur substitution of the 4-keto oxygen of uracil has little effect; tautomeric patterns of the *N*₁ and *N*₃ protons are comparable, whereas the sulphur substitution lowers the *pK*_a of uracil by less than one unit to 8.4 (refs 15, 16). After four rounds of selection–amplification, ribozyme activity was readily detected (Fig. 3). Pool activity increased another 50,000-fold in response to seven additional selective rounds that included shorter incubation times, lower ⁴⁵Ura concentration and mutagenic amplification (Fig. 3).

Ribozymes that had undergone 11 rounds of selection were cloned, and 35 random clones were sequenced. A family of 25 related sequences generated by the mutation of a single ancestral sequence dominated the round-11 isolates and was designated family A (Fig. 4a). The remaining isolates represented two other families, family B (eight isolates) and family C (two isolates). Restriction analysis of PCR DNA from each round of selection¹⁷ indicated that these were the only three families of nucleotide-synthesizing ribozymes to emerge to detectable levels ($\geq 4\%$ of the population) during the entire course of the selection and evolution (data not shown).

Synthesis of authentic tethered ⁴⁵U was confirmed for representatives of each ribozyme family by nearest-neighbour analysis of end-labelled product (Fig. 5). The ribozyme-synthesized nucleotide precisely comigrated with the 4-thiouridine 3'-phosphate (⁴⁵Up) standard in all six separation systems tested: two polyacrylamide gel systems and four thin-layer chromatography (TLC) systems. An analysis utilizing a two-dimensional TLC system known to resolve the many different modified bases of ribosomal RNA¹⁸ is shown (Fig. 5).

Isolates from each family promoted nucleotide formation with apparent second-order rate constants (*k*_{cat}/*K*_m values) up to 10⁷ times greater than our upper bound on the uncatalysed reaction rate (Fig. 4). In attempts to detect the uncatalysed reactions, a radiolabelled pRpp-derivatized oligonucleotide was incubated with ⁴⁵Ura and the reaction mixture was resolved on APM gels. ⁴⁵U synthesis was not detected, even though these assays would have readily detected a reaction as slow as $6 \times 10^{-7} \text{ M}^{-1} \text{ min}^{-1}$. The rates of a family A isolate fit well to a Michaelis–Menten curve with an

apparent *K*_m of $28 \pm 4 \text{ mM}$ and a *k*_{cat} of $0.13 \pm 0.012 \text{ min}^{-1}$ (means $\pm$ s.e.m.; Fig. 4b), although it should be noted that solubility constraints prevented rate measurements with ⁴⁵Ura concentrations above 14 mM. Representatives from the other two sequence families had comparable *k*_{cat}/*K*_m values but did not begin to display saturable behaviour at soluble concentrations of ⁴⁵Ura (Fig. 4b), suggesting poorer binding to ⁴⁵Ura but faster catalysis on encountering ⁴⁵Ura.

All three ribozymes had high specificity for the ⁴⁵Ura substrate. No thio-containing product was detected on APM gels after body-labelled ribozyme-pRpp was incubated with any of six other thio-substituted pyrimidine bases (2-thiouracil, 2,4-thiouracil, 2-thiocytosine, 2-thiopyrimidine, 2-thiopyridine and 5-carboxy-2-thiouracil; limits of detection, 4×10^{-3} to $1 \times 10^{-4} \text{ M}^{-1} \text{ min}^{-1}$, depending on whether the product migrated to an area of the gel with high or low background). The family A isolate reacted very slowly with radiolabelled uracil, with a rate ($2 \times 10^{-4} \text{ M}^{-1} \text{ min}^{-1}$) comparable to that of the pool with ⁴⁵Ura after four rounds of selection.

Protein enzymes that synthesize pyrimidine nucleotides are thought to catalyse an S_N1 reaction by stabilizing an oxocarbenium at the C1' carbon of the reaction centre^{12,13}. One challenge for these enzymes is to avoid the hydrolysis of pRpp to ribose 5'-phosphate and pyrophosphate, which occurs if water rather than the pyrimidine base is proximal to the highly reactive carbocation. The enzyme that synthesizes orotidine (6-carboxyuridine) avoids hydrolysing pRpp by excluding water from the active site and by promoting carbocation formation only after a conformational change induced by binding of orotate (6-carboxyuracil)^{12,13}. As with the metabolic enzymes, our ribozymes were selected to avoid pRpp hydrolysis; any catalyst that hydrolysed its pRpp moiety before ⁴⁵Ura addition (for example, during the T4 RNA ligase incubation needed to attach the pRpp) would have been inactive for ⁴⁵U synthesis. Thus, we examined the degree to which the ribozymes promoted the hydrolysis of the tethered pRpp. Ribozymes from the three families promoted the hydrolysis of their pRpp moiety at rates 12–23 times faster than uncatalysed hydrolysis (Fig. 4a). Nevertheless their *k*_{cat} values for ⁴⁵U formation were ≥ 60 times faster than their rates of catalysed hydrolysis. Either RNA does have the ability to exploit highly reactive reaction intermediates or these ribozymes have employed a new strategy for promoting glycosidic bond formation, that is, by stabilizing a transition state with more S_N2 character.

As with many ribozymes and the metabolic enzymes that synthesize nucleotides, all three ribozyme families required divalent cations for activity. For each round of selection, Mg²⁺ (25 mM), Ca²⁺ (1 mM) and Mn²⁺ (0.5 mM) had been provided. However, Ca²⁺ was dispensable for all three families. Although the three families were active in the presence of Mn²⁺ as the only divalent cation, all preferred Mg²⁺ over Mn²⁺ as the major divalent cation. The family A isolate did not require Mn²⁺, with only a twofold decrease in activity observed in the absence of Mn²⁺. The family B and C ribozymes required Mn²⁺, with the activity in the presence of 25 mM Mg²⁺ reaching a plateau at 1 mM Mn²⁺. The family B ribozyme did not require Mn²⁺ for stimulating pRpp hydrolysis, suggesting that for this family Mn²⁺ has a role in the binding or proper orientation of the ⁴⁵Ura, consistent with the thiophilic nature of Mn²⁺ compared with Mg²⁺ and Ca²⁺ (ref. 19).

Ribozymes of an RNA world would have needed to promote numerous reactions involving small organic molecules²⁰. An important question in this regard is whether RNA can perform covalent chemistry with substrates smaller than purine nucleosides. Ribonucleosides as small as adenosine and 2-aminopurine (*M*_r 267.2) are substrates for self-splicing intron derivatives^{21,22}. ⁴⁵Ura (*M*_r 128.1) is half the size of these nucleoside substrates and within the size range of the smallest aptamer targets, valine and arginine²³ (*M*_r 117.2 and 174.2, respectively). The findings that a catalytic RNA

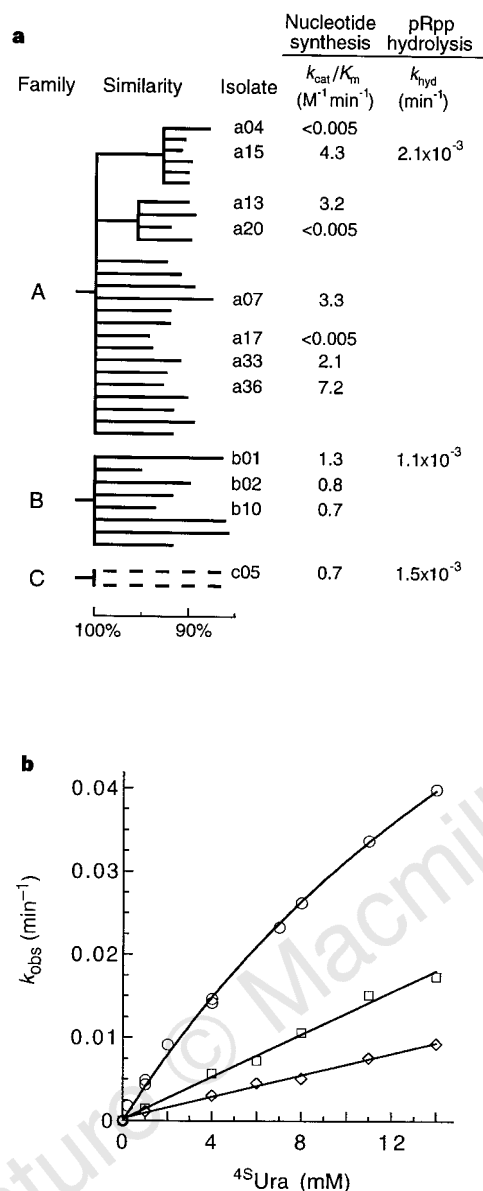


Figure 4 Analysis of ribozyme sequences and their catalytic proficiencies. **a**, Three families of ribozymes with rates of ^{45}U production and pRpp hydrolysis. Sequence analysis is summarized by family trees, with each branch terminus representing one of the 35 sequenced isolates. Similarity to the family consensus sequence varied from 94 to 86%, as indicated by the horizontal length of the branches. Family A has two dominant subfamilies, which presumably emerged through preferential expansion of superior early lineages. Because the family C consensus sequence could not be fully determined from only two isolates, similarity to the consensus is reported as a range (dashed lines). **b**, Rate of nucleotide formation (k_{obs}) as a function of ^{45}Ura concentration for isolates a15 (circles), b01 (squares) and c05 (diamonds). The rates of isolates b01 and c05 fit well to linear functions indicating k_{cat}/K_m values of 1.29 ± 0.03 and $0.67 \pm 0.02 M^{-1} min^{-1}$, respectively. The line shown for isolate a15 is the non-linear least-squares fit of a Michaelis-Menten curve to the data and suggests an apparent K_m of $28 \pm 4 mM$ and a k_{cat} of $0.13 \pm 0.012 min^{-1}$ (mean $\pm$ s.e.m.), although these values must be viewed with caution because solubility constraints prevented the examination of ^{45}Ura concentrations above 14 mM. For example, we cannot discount the possibility that ^{45}Ura was beginning to occupy an inhibitory site rather than the catalytic site. The linear behaviour of the other two isolates suggests that ^{45}Ura within this concentration range does not aggregate or affect metal-ion availability.

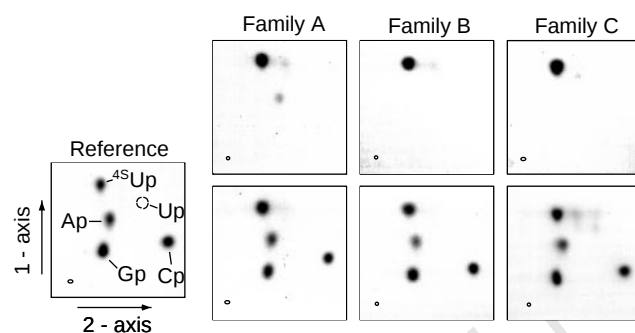


Figure 5 Two-dimensional TLC analysis¹⁸ of ribozyme-synthesized ^{45}U after labelling and digestion with nuclease. Ribozyme product RNA was extended by one nucleotide by using α - ^{32}P -cordycepin triphosphate (3'-deoxyATP) and poly(A) polymerase, then purified on an APM gel. Ribonuclease T2 digestion reduced all the labelled material into nucleoside 3'-phosphates, with the labelled phosphate residing on the ribozyme-synthesized nucleotide. Digests were spotted on 10 $\times$ 10 cm cellulose TLC plates (Baker flex) presoaked in 1:10 saturated- $(NH_4)_2SO_4:H_2O$. The first axis was run in 80% v/v ethanol, and after air-drying for 5 min, the second axis of 40:1 saturated- $(NH_4)_2SO_4$:isopropanol was run. The reference panel shows the migration of ^{45}Up , Up, Ap, Cp and Gp. Unlabelled or body-labelled ^{45}U -containing RNA was included as carrier in the digestion (top and bottom panels, respectively). Carrier RNA was generated by transcription, with ^{45}UTP replacing UTP.

can specifically recognize and utilize such a substrate and that RNA can efficiently promote the chemistry of glycosidic bond formation support the prospects of ribozyme-based metabolic pathways in the RNA world. Another important step would be to generate catalytic sequences capable of using not one but two small-molecule substrates. With regard to this goal it is encouraging that after optimization by evolution and engineering *in vitro*, a ribozyme motif initially selected on the basis of a reaction using an attached nucleoside triphosphate was able to promote a reaction using free nucleoside triphosphates²⁴. Similarly, our new ribozymes offer a basis for developing catalysts that synthesize ^{45}U from two small molecules, ^{45}Ura and pRpp. □

Methods

Pool construction and amplification. The double-stranded DNA pool with sequence *TTCTAATACGACTCACTATAGGACCGAGAAGTTACCC-N₇₆-CCTTGG-N₇₆-GGCACC-N₇₆-ACGCACATCGCAGCAAC* (italics, T7 promoter; -N₇₆, 76-nucleotide random-sequence segment) was constructed starting with three synthetic single-stranded pools, as described previously^{17,25}. The phosphoramidite ratio for random-sequence DNA synthesis was normalized to account for differing coupling rates (0.26:0.25:0.29:0.20 dA:dC:dG:dT molar ratio). Sequencing 2,030 random-sequence positions from arbitrary clones verified that the nucleotide composition was nearly equal (529:519:482:500 A:C:G:T). The DNA pool was transcribed *in vitro* and 16.4 nmol RNA (six copies, on average, of each pool sequence) were used for the first round of selection.

Tethering pRpp and $p^{45}U$ to RNA. pRpp was linked to RNA by exploiting the finding that specificity for the donor substrate of T4 RNA ligase can be bypassed by pre-adenylation²⁶ (Fig. 2a). Adenylylated pRpp (AppRpp) was synthesized by reacting 300 mM adenosine 5'-phosphorimidazole²⁷ with 600 mM pRpp in 200 mM $MgCl_2$ for 2 h at 50 °C. After ligation (2–4 μM gel-purified RNA, 65 μM HPLC-purified AppRpp, 50 mM HEPES pH 8.3, 10 mM $MgCl_2$, 3.3 mM dithiothreitol (DTT), 10 $\mu g/ml$ BSA, 8.3% v/v glycerol, 1 U μl^{-1} enzyme from Pharmacia, for 4 h at room temperature) and extraction with phenol-chloroform, RNA was recovered by precipitation with ethanol. The efficacy of ligation was verified by using matrix-assisted mass spectrometry of a 9-nt ligation product before and after treatment with alkaline phosphatase, which removed the 1' pyrophosphate. At each round, the efficiency of ligation

to the pool RNA was determined by the gel mobility of 3'-terminal fragments cleaved by a DNA enzyme²⁸ targeted to the 3' constant segment. RNAs extended with ⁴⁵U (for use as synthetic standards) were generated, using App⁴⁵U instead of AppRpp.

Ribozyme selection. ⁴⁵Ura was synthesized²⁹ and further purified by reverse-phase HPLC. For each round of selection the RNA-pRpp pool was incubated with ⁴⁵Ura (Fig. 2b) using the concentrations and incubation times outlined in Fig. 3 ($\leq 0.3 \mu\text{M}$ pool RNA, 50 mM Tris-HCl pH 7.5, 150 mM KCl, 25 mM MgCl₂, 1 mM CaCl₂, 0.5 mM MnCl₂ at 23 °C). Ribozyme reactions were stopped by the addition of one volume of gel loading buffer (90% formamide, 50 mM EDTA). RNA with a 3'-terminal ⁴⁵U was resolved from other species on denaturing APM gels^{14,30} (8 M urea/5% polyacrylamide gels, cast with 80 μM N-acryloylaminophenylmercuric acetate). During each round of selection, the ribozyme reaction was split in half. One half contained radiolabelled pool RNA that facilitated the detection and purification of emergent ribozymes; the other half contained unlabelled pool and a trace amount of radiolabelled synthetic standard (an RNA pool with a 3'-terminal ⁴⁵U but with primer-binding sequences incompatible with reverse transcription and PCR) that was used to locate and monitor the recovery of reacted RNA. The gel fragment containing RNA-⁴⁵U was excised (Fig. 2c), and RNA was eluted in 300 mM NaCl, 20 mM DTT, then precipitated with ethanol. After round 1, RNA containing ⁴⁵U was further purified on a second mercury gel and then biotinylated (Fig. 2d) by resuspending precipitated RNA in 25 mM potassium phosphate pH 8.4, 3 mM iodoacetyl-LC-biotin (Pierce), 50% v/v dimethylformamide. After 3 h at room temperature in the dark, the reaction was quenched with DTT and diluted 10-fold; RNA was then twice precipitated with ethanol to remove excess biotin. About 75% of material end-labelled with ⁴⁵U was biotinylated, as judged by a streptavidin gel-shift assay. Capture of biotinylated RNA, reverse transcription, PCR amplification and transcription *in vitro* were as described previously²⁵.

Ribozyme assays. Unless stated otherwise, ribozyme isolates were assayed under conditions similar to those of the selection (0.5 μM ³²P-labelled ribozyme RNA, 50 mM N,N-bis(2-hydroxyethyl)-2-aminoethanesulphonic acid (BES) pH 7.5, 150 mM KCl, 25 mM MgCl₂, 0.5 mM MnCl₂ at 23 °C). The rate of ⁴⁵U synthesis (k_{obs}) at a given ⁴⁵Ura concentration was determined by the best fit of k and β to the equation $F = \beta(1 - e^{-kt})$, where F is the fraction reacted (ascertained by phosphorimaging of the APM gel), $k = k_{\text{obs}} + k_{\text{hyd}}$, $\beta = F_{\text{max}}k_{\text{obs}}/(k_{\text{obs}} + k_{\text{hyd}})$, k_{hyd} is the rate of pRpp hydrolysis, and t is time. F_{max} , the maximal active fraction (typically 0.15–0.20) was determined by using time courses with ⁴⁵Ura in excess of 4 mM. Factors contributing to the low F_{max} values included: incomplete pRpp ligation (lowering F_{max} by 10–30%), 3' heterogeneity from untemplated residues added during transcription (lowering F_{max} by 40–50%), and ribozyme misfolding. The k_{hyd} values were determined by varying ⁴⁵Ura concentration and observing differences in the asymptotic fraction reacted. The k_{hyd} values were confirmed independently by monitoring the inactivation of RNA-pRpp in the absence of ⁴⁵Ura. The rate of uridine synthesis was determined by using isolate a15 and a random-sequence control RNA, both activated with pRpp. Each RNA (7 μM) was incubated with 0.5 mCi [5,6-³H]uracil (NEN) in standard buffer conditions. At 0, 2, 4 and 8 h, 160 μl aliquots were quenched by the addition of EDTA and unlabelled uracil. RNA was filtered (Centricon spin filters; Amicon), purified on a polyacrylamide gel and precipitated with ethanol. Uracil counts associated with the RNA were determined by scintillation (Formula-989 fluid, Packard) and corrected for RNA recovery. A similar approach with [2-¹⁴C]uracil indicated a comparable rate (within twofold). Additional analysis of RNA recovered from the scintillation fluid (precipitation with NaCl and six volumes of ethanol, and resuspension with unlabelled uridine) confirmed that the incorporated counts were from 3'-terminal uridine synthesis; RNA was base-hydrolysed and nucleotides were separated by reverse-phase HPLC and scintillation-counted. The uncatalysed reaction rates were examined with a 9-nt RNA-pRpp conjugate that was ³²P-labelled at its 5' terminus. After incubation with 6.4 mM ⁴⁵Ura for 4 days in the buffer used for the selection, no RNA-p⁴⁵U product was observed on APM gels. A single gel would have readily detected uncatalysed RNA-p⁴⁵U formation with a rate $\geq 2 \times 10^{-6} \text{ M}^{-1} \text{ min}^{-1}$ (correcting for the RNA-pRpp lost to hydrolysis), whereas a serial APM gel analysis lowered detection limits to $6 \times 10^{-7} \text{ M}^{-1} \text{ min}^{-1}$. The RNA-pRpp hydrolysed during the 4-day time course indicated an uncatalysed k_{hyd} of $9 \times 10^{-5} \text{ min}^{-1}$.

Received 13 March; accepted 20 July 1998.

- Joyce, G. F. & Orgel, L. E. in *The RNA World* (eds Gesteland, R. F. & Atkins, J. F.) 1–25 (Cold Spring Harbor Lab., Cold Spring Harbor, NY, 1993).
- Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis VI. Synthesis of purine nucleosides. *J. Mol. Biol.* **67**, 25–33 (1972).
- Fuller, W. D., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis VII. Solid-state synthesis of purine nucleosides. *J. Mol. Evol.* **1**, 249–257 (1972).
- Larralde, R., Robertson, M. P. & Miller, S. L. Rates of decomposition of ribose and other sugars: implications for chemical evolution. *Proc. Natl Acad. Sci. USA* **92**, 8158–8160 (1995).
- Mizuno, T. & Weiss, A. H. Synthesis and utilization of formose sugars. *Adv. Carbohydr. Chem. Biochem.* **29**, 173–227 (1974).
- Pitsch, S., Eschenmoser, A., Gedulin, B., Hui, S. & Arrhenius, G. Mineral induced formation of sugar phosphates. *Origin Life Evol. Biosphere* **25**, 297–334 (1995).
- Oro, J. Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive earth conditions. *Nature* **191**, 1193–1194 (1961).
- Sanchez, R. A., Ferris, J. P. & Orgel, L. E. Studies in prebiotic synthesis. II. Synthesis of purine precursors and amino acids from aqueous hydrogen cyanide. *J. Mol. Biol.* **30**, 223–253 (1967).
- Ferris, J. P., Sanchez, R. A. & Orgel, L. E. Studies in prebiotic synthesis. 3. Synthesis of pyrimidines from cyanoacetylene and cyanate. *J. Mol. Biol.* **33**, 693–704 (1968).
- Stoks, P. G. & Schwartz, A. W. Uracil in carbonaceous meteorites. *Nature* **282**, 709–710 (1979).
- Robertson, M. P. & Miller, S. L. An efficient prebiotic synthesis of cytosine and uracil. *Nature* **375**, 772–774 (1995).
- Bhatia, M. B., Vinitsky, A. & Grubmeyer, C. Kinetic mechanism of orotate phosphoribosyltransferase from *Salmonella typhimurium*. *Biochemistry* **29**, 10480–10487 (1990).
- Tao, W., Grubmeyer, C. & Blanchard, J. S. Transition state structure of *Salmonella typhimurium* orotate phosphoribosyltransferase. *Biochemistry* **35**, 14–21 (1996).
- Igloi, G. L. Interaction of tRNAs and of phosphorothioate-substituted nucleic acids with an organomercurial. Probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry* **27**, 3842–3849 (1988).
- Wierzchowski, K. L., Litonska, E. & Shugar, D. Infrared and ultraviolet studies on the tautomeric equilibria in aqueous medium between monoanionic species of uracil, thymine, 5-fluorouracil, and other 2,4-diketopyrimidines. *J. Am. Chem. Soc.* **87**, 4621–4629 (1965).
- Psoda, A., Kazimierzczuk, Z. & Shugar, D. Structure and tautomerism of the neutral and monoanionic forms of 4-thiouracil derivatives. *J. Am. Chem. Soc.* **96**, 6832–6839 (1974).
- Bartel, D. P. & Szostak, J. W. Isolation of new ribozymes from a large pool of random sequences. *Science* **261**, 1411–1418 (1993).
- Gray, M. W. The presence of 2'-O-methylpseudouridine in the 18S + 26S ribosomal ribonucleates of wheat embryo. *Biochemistry* **13**, 5453–5463 (1974).
- Jaffe, E. K. & Cohn, M. Diastereomers of the nucleoside phosphorothioates as probes of the structure of the metal nucleotide substrates and of the nucleotide binding site of yeast hexokinase. *J. Biol. Chem.* **254**, 10839–10845 (1979).
- Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl Acad. Sci. USA* **86**, 7054–7058 (1989).
- Michel, F., Hanna, M., Green, R., Bartel, D. P. & Szostak, J. W. The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* **342**, 391–395 (1989).
- Been, M. D. & Perrotta, A. T. Group I intron self-splicing with adenosine: evidence for a single nucleoside-binding site. *Science* **252**, 434–437 (1991).
- Gold, L., Polisky, B., Uhlenbeck, O. & Yarus, M. Diversity of oligonucleotide functions. *Annu. Rev. Biochem.* **64**, 763–797 (1995).
- Eklund, E. H. & Bartel, D. P. RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373–376 (1996).
- Eklund, E. H. & Bartel, D. P. The secondary structure and sequence optimization of an RNA ligase ribozyme. *Nucleic Acids Res.* **23**, 3231–3238 (1995).
- England, T. E., Gumpert, R. I. & Uhlenbeck, O. C. Dinucleoside pyrophosphates are substrates for T4-induced RNA ligase. *Proc. Natl Acad. Sci. USA* **74**, 4839–4842 (1977).
- Lohrmann, R. & Orgel, L. E. Preferential formation of (2'-5')-linked internucleotide bonds in non-enzymatic reactions. *Tetrahedron* **34**, 853–855 (1978).
- Santoro, S. W. & Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl Acad. Sci. USA* **94**, 4262–4266 (1997).
- Mizuno, Y., Ikehara, M. & Watanabe, K. A. Potential antitumor agents. I. Selective thiation of uracil and 1,2,4-triazine-3,5(2H,4H)-dione (6-azauracil). *Chem. Pharmac. Bull.* **10**, 647–652 (1962).
- Wecker, M., Smith, D. & Gold, L. *In vitro* selection of a novel catalytic RNA: characterization of a sulfur alkylation reaction and interaction with a small peptide. *RNA* **2**, 982–994 (1996).

Acknowledgements. We thank J. Stubbe, P. Zamore and members of the lab for helpful comments on the manuscript, and G. Joyce for providing the sequence of the RNA-cleaving DNA enzyme²⁸ before publication. This work was supported by an MRC (Canada) postdoctoral fellowship to P.J.U. and a grant from the Searle Scholars Program/The Chicago Community Trust to D.P.B.

Correspondence and requests for materials should be addressed to D.P.B. The nine active sequences have been deposited in GenBank under accession numbers AF051883–51891.

Deep-ocean gradients in the concentration of dissolved organic carbon

Dennis A. Hansell & Craig A. Carlson

Bermuda Biological Station for Research, Inc., St Georges, GE01, Bermuda

There is as much carbon in dissolved organic material in the oceans as there is CO₂ in the atmosphere¹, but the role of dissolved organic carbon (DOC) in the global carbon cycle is poorly understood. DOC in the deep ocean has long been considered to be uniformly distributed^{2,3} and hence largely refractory to biological

Table 1 Deep ocean concentrations of dissolved organic carbon

Location	Position	Depth range (m)	DOC (μM)	<i>N</i>	s.d.	Comments
North Atlantic	75° N, 0° W	1,250–3,650	48.1	10	0.3	2 vertical profiles, April 1995
	48° N; 13°, 41° and 48° W	1,100–4,800	45.1	19	0.4	Mean of 3 stations, 1997
	32° N, 164° W	1,000–4,000	43.6	35	0.3	Mean of 6 profiles, 1997
Circumpolar Current	60° S, 170° W	1,000–4,000	41.5	7	0.4	Single profile, 1997*
	61° S, 170° W	2,000	41.9	33	0.3	Mean of replicate samples, 1996*
Ross Sea	75–77° S, 178° E–178° W	0–700	41.7	241	0.3	Regional mean, late winter 1996*
Indian Ocean	36° S, 95° E	1,000–2,000	43.3	3	0.1	Single profile, 1995
	32° S, 80° E	1,500–2,000	42.9	2	–	Single profile, 1995
	26° S, 80° E	1,000–2,000	43.3	3	0.0	Single profile, 1995
	14° S, 80° E	1,000–4,000	42.9	6	0.7	Single profile, 1995
	8° S, 80° E	1,500–5,500	43.6	3	0.3	Single profile, 1995
	3° N, 80° E	1,500–3,000	43.3	4	0.2	Single profile, 1995
Arabian Sea	13–20° N, 60–65° E	1,000–4,300	42.8	27	0.8	Mean of 8 stations, 1995*
Pacific Ocean	56.5° S, 170° W	1,000–5,000	42.3	5	0.3	Single profile, 1996
	52.5° S, 170° W	1,000–5,000	42.3	5	0.1	Single profile, 1996
	33.5° S, 170° W	1,000–5,000	43.0	5	0.1	Single profile, 1996
	0° S, 170° W	1,000–5,000	42.7	4	0.1	Single profile, 1996
	22.75° N, 158° W	900–4,750	38.8	9	0.9	Single profile, 1997
	58° N, 148° W	1,000–1,500	33.8	3	0.4	Single profile, 1997

Details on the sampling and results from the deep ocean survey of DOC concentrations. All samples were collected with Niskin bottles on CTD rosettes. DOC is the mean of the concentrations in one or more vertical profiles at a given site; *N* is the number of depths sampled over the depth range indicated; s.d. is the standard deviation of the means from the several depths at each site; the dates indicate the year of sampling. Samples from the South Atlantic Ocean remain lacking in this survey. Samples selected for calculation of the mean were those from the zone in the water column (generally >1,000 m) where vertical gradients in DOC concentration were not evident. Bottom depth in the Ross Sea is generally <1,000 m, so shallower depths were taken as no vertical gradient was evident.

* These analyses were performed at sea. All other analyses were done over a period of a few months in the shore-based laboratory. Samples not run at sea were stored at –20°C in pre-cleaned glass vials or polyethylene bottles. These storage techniques have been thoroughly tested over the past several years and have demonstrated no artefactual results on the micromolar scale. Deep Sargasso Sea reference water (2,600 m; DOC, $43.6 \pm 0.2 \mu\text{M}$) was acidified with H_3PO_4 , ampouled in glass and stored unfrozen. Experience with this reference material has shown that it remains stable for a minimum of several years. While in the field, deep Sargasso Sea reference water was evaluated every fourth to sixth analysis. While in the shore laboratory, the reference material was run every second and third analysis. High-precision analyses for DOC concentrations were done using a non-commercial high-temperature combustion system with procedures and blank corrections as described²⁹. Analytical precision in this work on replicated samples was approximately 0.3 μM . Samples were not filtered, but because particulate organic carbon loads (organic carbon retained on a GF/F filter) are generally less than a few tenths of micromolar in the deep ocean, the data are reported as DOC.

decay⁴. But the turnover of DOC, and therefore its contribution to the carbon cycle, has been evident from radiocarbon dating studies^{5,6}. Here we report the results of a global survey of deep-ocean DOC concentrations, including the region of deep-water formation in the North Atlantic Ocean, the Circumpolar Current of the Southern Ocean, and the Indian and Pacific oceans. DOC concentrations decreased by 14 micromolar from the northern North Atlantic Ocean to the northern North Pacific Ocean, representing a 29% reduction in concentration. We evaluate the spatial patterns in terms of source/sink processes. Inputs of DOC to the deep ocean are identifiable in the mid-latitudes of the Southern Hemisphere, but the mechanisms have not been identified with certainty.

Deep water samples were collected from numerous sites in the global ocean (Fig. 1, Table 1). Measured concentrations of DOC were determined relative to deep Sargasso Sea reference water (see Table 1 for methods). Concentrations were highest in the Greenland Sea ($48 \pm 0.3 \mu\text{M}$ C), decreasing to 45.1 ± 0.4 and $43.6 \pm 0.3 \mu\text{M}$ near 48° N and 32° N, respectively, in the North Atlantic (Fig. 2). Concentrations in the Circumpolar Current (61° S) and the Ross Sea (75° S) were 41.5–41.9 μM , then increased to the north in the central Indian Ocean and Arabian Sea (42.8–43.6 μM C), as well as in the western South Pacific Ocean (42.3–43.0 μM C). Concentrations decreased into the North Pacific, where the global low of $33.8 \pm 0.4 \mu\text{M}$ was found at the northernmost site sampled. Previously published data sets were not used in our synthesis because a reference material common to that used here was not employed.

DOC in the deep ocean comprises a large fraction of very old material (6,000 yr) and a smaller fraction of young material^{5,6}. The changes seen in DOC concentration along the deep path of the global 'conveyor belt'⁷ result from: the addition of young material at the sites of deep-water formation (particularly in the North Atlantic) and its microbial consumption at depth; isopycnal and diapycnal mixing of waters with various DOC loads; and small additions (solubilization)⁸ and losses (adsorption)⁹ as a result of transformation of particulate matter. That the large gradient seen in Fig. 1 was created by seasonal variability in deep DOC, perhaps linked to seasonality in the vertical flux of biogenic particles, is discounted by

the lack of short-term variability in the deep Sargasso and Arabian seas (Table 2). These two sites show strong seasonality in vertical particle flux, with flux being particularly high in the spring in the Sargasso Sea¹⁰ and in late summer in the Arabian Sea¹¹, but little variability in deep DOC over those periods.

The Greenland Sea samples are our best representation for concentrations of DOC in the source water for North Atlantic Deep Water (NADW) formation. NADW overrides and entrains Antarctic Bottom Water (AABW), of a largely Weddell Sea source¹², during its southward flow in the Atlantic. Assuming that the DOC concentrations reported here for the deep Circumpolar Current in the Pacific sector (170° W; Table 1) are the same as concentrations in AABW from the Weddell Sea (an assumption also made for the remaining discussion), then the mixing of these two source waters, microbial degradation, and other loss terms will result in a DOC

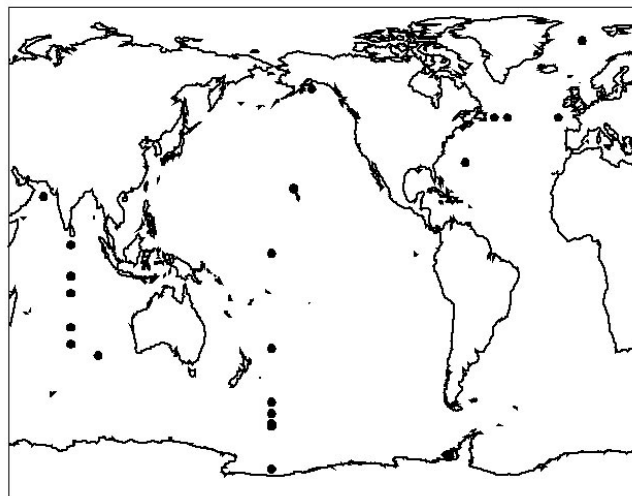


Figure 1 World map depicting locations sampled for the survey. Station locations are given in Table 1.

Table 2 Monthly variability in deep ocean DOC concentrations at two sites

Location	Period	DOC* (μM)	s.d.	N
Sargasso Sea	January	43.2	0.8	5
	February	43.5	0.2	7
	March	43.4	0.3	7
	April	43.4	–	2
	May	43.4	–	2
	June	43.6	0.5	7
	July	43.1	0.2	7
	August	43.5	0.7	6
	September	43.6	0.2	6
	October	43.4	0.3	7
	November	43.7	0.2	6
	December	43.7	0.2	6
Arabian Sea	July/August	42.5	0.8	97
	August/September	42.6	0.5	19
	November	42.9	0.8	85
	December	42.4	0.8	119

* Monthly variability in deep ocean DOC concentrations at two sites: the Sargasso Sea (32°N , 164°W ; sampled in 1996) and the Arabian Sea (in the area bounded by $13\text{--}20^{\circ}\text{N}$, $60\text{--}65^{\circ}\text{E}$; sampled in 1995). The Sargasso Sea values result from the monthly visits made to the site of the US JGOFS Bermuda Atlantic time-series study. The analyses were done approximately every month. The mean values reported for the Arabian Sea are for all deep-water analyses performed during US JGOFS Arabian Sea Process Cruises 4, 5, 6 and 7, respectively. These cruises coincided with the hydrologically defined southwest Monsoon (July and August), autumn intermonsoon (November) and the northeast Monsoon (December). "Period" indicates the nominal month of sampling; N, number of depths sampled; s.d., standard deviation about the mean for the samples taken from N depths. Sampling depths ranged from 900 to $>4,000\text{ m}$. Arabian Sea Cruise 4 samples were analysed in the laboratory of E. Peltzer; all others were done in the Hansell laboratory. Note, there is low temporal variability relative to the gradients of deep ocean DOC shown in Table 1.

concentration gradient in the Atlantic Ocean, decreasing from north to south. Deep-water silicate concentrations in the Atlantic¹³, useful for indicating the degree of mixing between the two components¹², suggest a dilution of NADW with $\sim 10\%$ AABW at the 32°N site. Such a dilution would account for a $0.6\text{ }\mu\text{M}$ decrease in DOC from Greenland Sea concentrations, so the remaining $3.8\text{ }\mu\text{M}$ decrease can be assigned to microbial consumption if it is the dominant loss term. The net rate of microbial utilization of DOC during the $\sim 80\text{-yr}$ transit⁷ from the site of NADW formation to the 32°N sampling site was $0.05\text{ }\mu\text{M}\text{Cyr}^{-1}$, or half the total organic carbon oxidation (sum of DOC and sinking biogenic carbon oxidation) estimated from apparent oxygen utilization rates in the deep central Atlantic¹⁴. Total DOC mineralization along the path of deep flow from the site of

formation to 32°N is estimated at $0.64 \times 10^{12}\text{ g C yr}^{-1}$ (given 14 Sv (where $1\text{ Sv} = 10^6\text{ m}^3\text{ s}^{-1}$) of NADW flow¹⁵).

Estimates of microbial utilization of DOC employing concentration gradients thought to develop over long distances and a multi-decadal period requires two primary assumptions: that concentrations of DOC reported here for the site of deep-water formation are constant, such that measurements made there now adequately reflect concentrations several decades earlier; and that neither deep winter convection in the Labrador Sea¹⁶ nor diapycnal mixing of intermediate water of a Labrador Sea source alter deep DOC concentrations significantly. These assumptions are not testable with the current data set.

Deep water in the Indian and Pacific oceans have as their source AABW and NADW^{17–19}. In the Indian Ocean, the Weddell Sea contributes to the densest water found in the western basins whereas the Ross Sea contributes to the eastern basins. The sampling conducted here (mostly along 80°E) was largely in the return flow of these source waters from the northern Indian Ocean. The raised concentration of DOC in the return flow may result from the introduction of DOC-rich subsurface water to the deep layers in the Arabian Sea and the Bay of Bengal. Outflow from marginal seas such as the Red Sea and the Persian Gulf could be particularly important. The subsurface Persian Gulf outflow (at 250 m in the mouth of the Gulf of Oman) has DOC concentrations in excess of $50\text{ }\mu\text{M}$ (ref. 20). It is likely that the Red Sea outflow, which reaches depths of $1,000\text{ m}$ in the Arabian Sea and has been reported as far south as 30°S (ref. 21), has similarly raised DOC concentrations. Alternatively, NADW may add DOC to the central Indian Ocean. NADW contributes $25 \pm 10\%$ of input to the deep Indian Ocean¹², so the increase in DOC observed along 80°E could be ascribed to NADW if concentrations in that water were $\sim 48.5\text{ }\mu\text{M}$ as it entered the Indian Ocean. This requirement is unlikely to be met, given the decrease in DOC along the path of NADW flow from the North Atlantic shown in Fig. 2. If the northern Indian Ocean is the source of the new DOC, then an estimate of carbon export can be made. 15 Sv of return flow from the deep Indian Ocean to the circumpolar deep water¹⁹, given the addition of $1.5\text{ }\mu\text{M}$ DOC, carries with it an additional carbon load of $8.5 \times 10^{12}\text{ g DOC yr}^{-1}$.

Mechanisms for the DOC concentration increase seen from the deep Circumpolar Current north into the South Pacific are more difficult to identify, but entrainment of DOC-rich overlying waters may contribute. Geostrophic currents along the section sampled (170°W) are largely southward at depths of $<3,000\text{ m}$, and largely northward at depths $>4,000\text{ m}$ (ref. 22). At intermediate depths, laterally adjacent currents flow in opposing directions. The south flowing waters are part of the subtropical gyre and equatorial upwelling systems, which have raised DOC concentrations above and in the main thermocline²³, whereas the northward flowing water is circumpolar deep water transiting the Samoan Passage into the deep equatorial and North Pacific region²⁴. Shear between these current systems will enhance the exchange of constituents such as DOC, but exact mechanisms and locations of transfer are uncertain.

Deep water concentrations of DOC south (33.5°S) and north (Equator) of the Samoan Passage (near 10°S , 170°W) were very similar (Fig. 2), whereas the concentration decreased rather abruptly into the North Pacific. The deep North Pacific contains the world's oldest water relative to time since ventilation with the atmosphere²⁵, with inflow from the Samoan Passage as its principal source^{19,24}. The very low DOC concentrations are likely to be due to microbial decomposition in excess of new carbon introduced by vertical entrainment or sinking particle solubilization. Additional loss of DOC in deep North Pacific waters has been proposed to occur by adsorption onto suspended particulate organic carbon⁹, which could in turn fall from the water column. This process could take on particular significance over the timescales of deep ocean circulation.

The total organic carbon stock in the global ocean estimated from

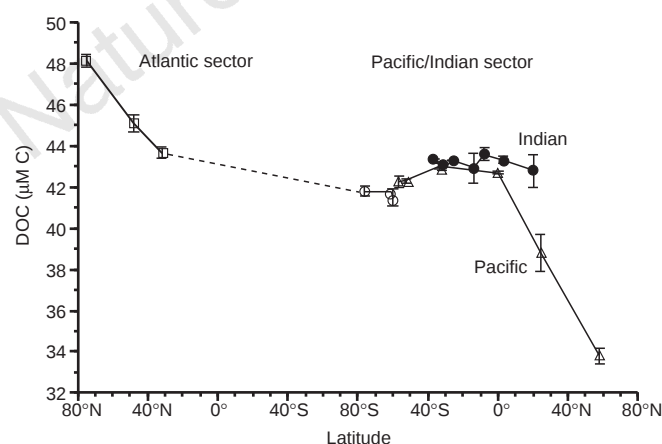


Figure 2 Mean concentrations of dissolved organic carbon (DOC; $\mu\text{M}\text{C}$) in the deep ocean. The means shown are of several depths at various sites in the global ocean (see Table 1 for locations and depth ranges sampled). The standard deviations of the means are shown as vertical bars through the data points. Solid lines connect data within ocean basins; the dashed line is used to demonstrate uncertainty in the South Atlantic Ocean as samples were not available for this analysis. Open boxes indicate data from the Atlantic Ocean, open circles the Southern Ocean, filled circles the Indian Ocean (including the Arabian Sea), and open triangles the Pacific Ocean.

the data in Table 1 and from ocean basin volumes is $\sim 685 \times 10^{15}$ g (57×10^{15} mol C), falling on the low side of existing estimates^{3,5,26}. Given the size of the oceanic reservoir of organic carbon, the contribution of this pool to the global carbon cycle remains surprisingly poorly resolved. With the clear signals for DOC concentration gradients in the deep ocean, large transformations and transport of organic carbon are evident. Concentration differences between the North Atlantic and the North Pacific result from differences in the thermohaline circulation patterns in those basins. The deep North Atlantic Ocean is filled with water relatively recently at the surface, with strong contributions of DOC-enriched subtropical gyre water via the North Atlantic Current. The deep North Pacific, on the other hand, is flushed by circumpolar deep water of an Antarctic source, with quite low initial concentrations of DOC (Fig. 2).

Further speculation on the biological and physical processes forcing the deep-ocean DOC signals will require a much larger and spatially complete survey of deep and intermediate ocean DOC concentrations. Regions of focus must include the Arctic Ocean (where increased DOC concentrations have been reported²⁷), the Weddell Sea (primary site of AABW formation), and the major sites of mode and intermediate water formation. We have assumed throughout this discussion that deep Weddell Sea water has DOC concentrations very similar to those of the measured Circumpolar and Ross Sea waters in the Pacific sector. Such may be a reasonable assumption, given the uniform distribution of other variables, such as $\Delta^{14}\text{C}-\text{CO}_2$, in the circumpolar waters²⁵. Higher concentrations of organic carbon have been reported, however, in the deep Weddell Sea²⁸, but comparison between data sets is not possible because a common analytical reference material was not used. These uncertainties highlight the need for a broadened and fully referenced survey of the deep ocean in order to identify the role of marine DOC in the oceanic carbon cycle. □

Received 26 January; accepted 30 June 1998.

- Hedges, J. I. Global biogeochemical cycles: progress and problems. *Mar. Chem.* **39**, 67–93 (1992).
- Menzel, D. W. The role of *in situ* decomposition of organic matter on the concentrations of non-conservative properties in the sea. *Deep-Sea Res.* **17**, 751–764 (1970).
- Martin, J. H. & Fitzwater, S. E. Dissolved organic carbon in the Atlantic, Southern and Pacific Oceans. *Nature* **356**, 699–700 (1992).
- Barber, R. T. Dissolved organic carbon from deep waters resist microbial oxidation. *Nature* **220**, 274–275 (1968).
- Williams, P. M. & Druffel, E. R. M. Radiocarbon in dissolved organic matter in the central North Pacific Ocean. *Nature* **330**, 246–248 (1987).
- Bauer, J. E., Williams, P. M. & Druffel, E. R. M. ^{14}C activity of dissolved organic carbon fractions in north-central Pacific and Sargasso Sea. *Nature* **357**, 667–670 (1992).
- Broecker, W. S. The great ocean conveyor. *Oceanography* **4**, 79–89 (1991).
- Smith, D., Simon, M., Alldredge, A. L. & Azam, F. Intense hydrolytic enzyme activity on marine aggregates and implications for rapid particle dissolution. *Nature* **359**, 139–142 (1992).
- Druffel, E. R. M., Bauer, J. E., Williams, P. M., Griffin, S. & Wolgast, D. Seasonal variability of particulate organic radiocarbon in the northeast Pacific Ocean. *J. Geophys. Res.* **101**, 20543–20552 (1996).
- Deuser, W. G. in *Scope 57 Particle Flux in the Ocean* (eds Ittekkot, V., Schafer, P., Honjo, S. & Depetris, P. J.) 185–198 (Wiley, New York, 1996).
- Honjo, S., Dymond, J., Prell, W. & Manganini, S. J. Monsoon controlled export fluxes to the interior of the Arabian Sea: JGOFS 1994–95 Experiment. *Deep-Sea Res.* **II** (submitted).
- Broecker, W. S. & Peng, T.-H. *Tracers in the Sea* (Lamont Doherty Geological Observatory, Columbia University, Palisades, NY, 1982).
- Knap, A. H. et al. *BATS Data Report B-5* (US JGOFS Planning and Coordination Office, Woods Hole Oceanographic Institution, MA, 1995).
- Broecker, W. S., Blanton, S., Smethie, W. M. & Ostlund, G. Radiocarbon decay and oxygen utilization in the deep Atlantic Ocean. *Global Biogeochem. Cycles* **5**, 87–117 (1991).
- Schmitz, W. J. Jr *On the World Ocean Circulation, I* (Woods Hole Oceanographic Institution Technical Report WHOI-96-03, MA, 1996).
- Sy, A. et al. Surprisingly rapid spreading of newly formed intermediate waters across the North Atlantic Ocean. *Nature* **386**, 675–679 (1997).
- Reid, J. L. & Lynn, R. J. On the influence of the Norwegian–Greenland and Weddell seas upon the bottom waters of the Indian and Pacific oceans. *Deep-Sea Res.* **18**, 1063–1088 (1971).
- Manly, A. W. & Reid, J. L. On the origins of deep and bottom waters of the Indian Ocean. *J. Geophys. Res.* **100**, 2417–2439 (1995).
- Schmitz, W. J. Jr *On the World Ocean Circulation, I* (Woods Hole Oceanographic Institution Technical Report WHOI-96-08, MA, 1996).
- Hansell, D. A. & Peltzer, E. T. Spatial and temporal variations of total organic carbon in the Arabian Sea. *Deep-Sea Res.* **II** (in the press).
- Dileep, K. M. & Li, Y.-H. Spreading of water masses and regeneration of silica and ^{226}Ra in the Indian Ocean. *Deep-Sea Res.* **II** **43**, 83–110 (1996).
- Reid, J. L. On the total geostrophic circulation of the South Pacific Ocean: Flow patterns, tracers and transports. *Prog. Oceanogr.* **16**, 1–61 (1986).
- Hansell, D. A. & Waterhouse, T. Y. Controls on the distribution of organic carbon and nitrogen in the eastern Pacific Ocean. *Deep-Sea Res.* **144**, 843–857 (1997).

- Roemmich, D., Hautala, S. & Rudnick, D. Northward abyssal transport through the Samoan passage and adjacent regions. *J. Geophys. Res.* **101**, 14039–14055 (1996).
- Stuiver, M., Quay, P. D. & Ostlund, H. G. Abyssal water carbon-14 distribution and the age of the world oceans. *Science* **219**, 849–851 (1983).
- Druffel, E. R. M., Williams, P. M., Bauer, J. E. & Ertel, J. R. Cycling of dissolved and particulate organic matter in the open ocean. *J. Geophys. Res.* **97**, 15630–15659 (1992).
- Anderson, L. G., Olsson, K. & Skoog, A. in *The Polar Oceans and Their Role in Shaping the Global Environment* (eds Johannessen, O. M., Muench, R. D. & Overland, J. E.) 255–262 (Geophys. Monogr. Ser., Am. Geophys. Union, Washington DC, 1994).
- Wedborg, M., Hoppema, M. & Skoog, A. On the relation between organic and inorganic carbon in the Weddell Sea. *J. Mar. Syst.* (in the press).
- Hansell, D. A., Carlson, C. A., Bates, N. & Poisson, A. A re-assessment of carbon export in the Equatorial Pacific Ocean. *Deep-Sea Res.* **II** **44**, 2115–2130 (1997).

Acknowledgements. We thank R. Parsons for her fine analytical work and T. Weingartner, D. Karl, D. Wallace and J. Urban-Rich for assistance during sample collections. This work was supported by the US NOAA and NSF.

Correspondence and requests for materials to D.A.H. (e-mail: dennis@bbsr.edu).

Plastic deformation of silicate spinel under the transition-zone conditions of the Earth's mantle

Shun-ichiro Karato*, Catherine Dupas-Bruzeck† & David C. Rubie†

* University of Minnesota, Department of Geology and Geophysics, Minneapolis, Minnesota 55455, USA

† Bayerisches Geoinstitut, Universität Bayreuth, D-95440, Germany

The dynamics of the Earth's deep interior are controlled to a large extent by rheological properties^{1,2}. Until recently, however, experimental studies on the rheological properties of materials thought to be present in the Earth's deep interior have been limited to relatively low pressures. Most previous estimates of rheology have therefore been based on either large extrapolations of low-pressure experimental data^{3,4} or inferences from geodynamical observations^{5–7}. Such studies have provided only weak constraints on the complicated rheological structure expected in the transition zone of the Earth's mantle (between 410 and 660 km depth) where a series of phase transformations occur in silicate minerals⁸. Here we report the results of a direct experimental study of deformation, under transition-zone conditions, of the spinel phase of $(\text{Mg,Fe})_2\text{SiO}_4$ (ringwoodite; thought to be present in the Earth's transition zone). Relatively coarse-grained samples show evidence of dislocation creep with dislocation structures similar to those observed in oxide and germanate spinels^{9,10}, which have significantly higher creep strengths than olivine^{10,11}. In contrast, a fine-grained sample shows evidence for grain-size-sensitive creep. These observations suggest that a ringwoodite-rich layer of the transition zone is likely to have a higher viscosity than the olivine-rich upper mantle³, whereas a subducting slab in the deep transition zone may lose its strength if significant grain-size reduction occurs^{12–14}.

Recent high-resolution seismic tomography^{15–17} and numerical modelling^{18–21} suggest that the interactions of convection currents

Table 1 Summary of experiments

Run	<i>P</i> (GPa)	<i>T</i> (K)	<i>t</i> (h)	<i>d</i> (μm)	<i>γ</i> (%)
no. 704	16	1,400	5	0.4 (S)* 3 (L)†	56
no. 710	16	1,600	5	6	±
no. 1740	16	1,600	10	5	102

Here *P* is pressure, *T* is temperature, *t* is time, *d* is grain-size, and *γ* is shear strain.

* Most of the sample. S indicates small grains.

† A few grains occupying <0.1% of the sample volume. L indicates large grains.

‡ Strain was not measured because the sample was cut at a wrong orientation.

with the transition zone of the mantle are complicated, causing the geometry of subducted slabs to be variable. These complications are mostly the consequence of changes in physical properties in the transition zone caused by phase transformations, particularly rheological properties. However, studies of the effects of phase transformations on rheological properties have been difficult, and no direct experimental studies have been performed under the conditions equivalent to the transition zone of the Earth except for a preliminary report on wadsleyite^{22,23} where deformation conditions were not well defined.

Here we report the results of high-pressure, high-temperature, large-strain shear deformation experiments performed on ringwoodite with different grain sizes. We used a newly developed experimental technique of high-pressure, high-temperature deformation²⁴. The starting material is a synthetic olivine aggregate with $\text{Fe}/(\text{Fe} + \text{Mg}) = 0.4$, synthesized from a mixture of oxides. About 1 wt% excess silica was added to ensure that the stoichiometry is on the silica-rich side. The sample was prepared by hot-pressing at 300 MPa and 1,473 K, and has low porosity (<2%). The average grain size is 5–10 μm and grain boundaries show an equilibrium geometry.

A thin slice (~ 0.2 mm thick) of the sample was sandwiched between two alumina pistons, the ends of which were cut at 45° with respect to their long axis. The pistons and sample were inserted in a gold tube that was placed in a multianvil octahedral pressure cell²⁴. The samples were first pressurized to 16 GPa in 5 h. Then the temperature was raised to a desired value in ~ 10 min. We choose 1,600 K ($T/T_m = 0.61$, where T is temperature and T_m is melting temperature; here T_m is²⁵ 2,600 K) and 1,400 K ($T/T_m = 0.54$). Under these conditions, $(\text{Mg,Fe})_2\text{SiO}_4$ olivine with $\text{Fe}/(\text{Fe} + \text{Mg}) = 0.4$ transforms to the spinel structure²⁶, and the grain size changes to an extent that depends on the conditions of transformation¹⁴. Samples were kept at a constant pressure and temperature for a certain period (2–10 h) to allow significant plastic deformation to develop (Table 1). After this period, temperature was quenched and the pressure was reduced over a period of 15 h.

After each experiment, the sample strain was measured from the rotation of a strain marker (Fig. 1). Most of the deformation occurred as simple shear and was homogeneous. A previous study²⁴ with a similar sample assembly showed that the deformation in this design occurs mostly at high pressures and temperatures and not during pressurization, and that the mode of deformation is predominantly stress relaxation at differential stresses of 1–3 GPa and average strain rates of 10^{-5} – 10^{-4} s^{-1} .

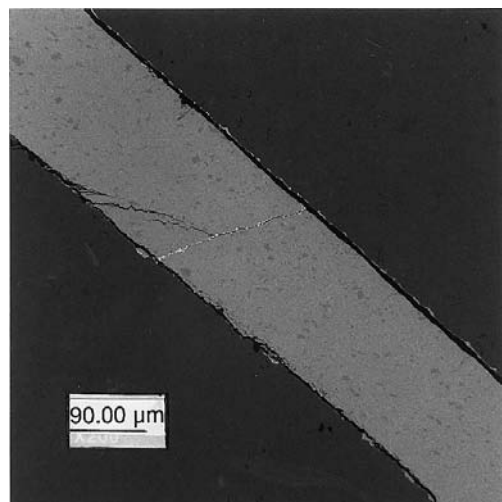


Figure 1 A scanning electron micrograph of a deformed specimen (no. 704). We note that the strain marker rotated from the original direction perpendicular to the sample–piston interface. The shear strain is estimated to be 56% and is nearly homogeneous.

Three deformed samples of ringwoodite have been investigated in detail using scanning electron microscopy (SEM), and two of these were investigated using transmission electron microscopy (TEM). One sample (no. 704) was transformed to spinel and deformed at 16 GPa and 1,400 K for 5 h. This sample has a small homogeneous grain-size of $\sim 0.5 \mu\text{m}$ with a small number (<0.1 vol.%) of relatively large grains ($\sim 3 \mu\text{m}$). Four observations are critical to identifying the mechanisms of deformation. (1) The TEM observations

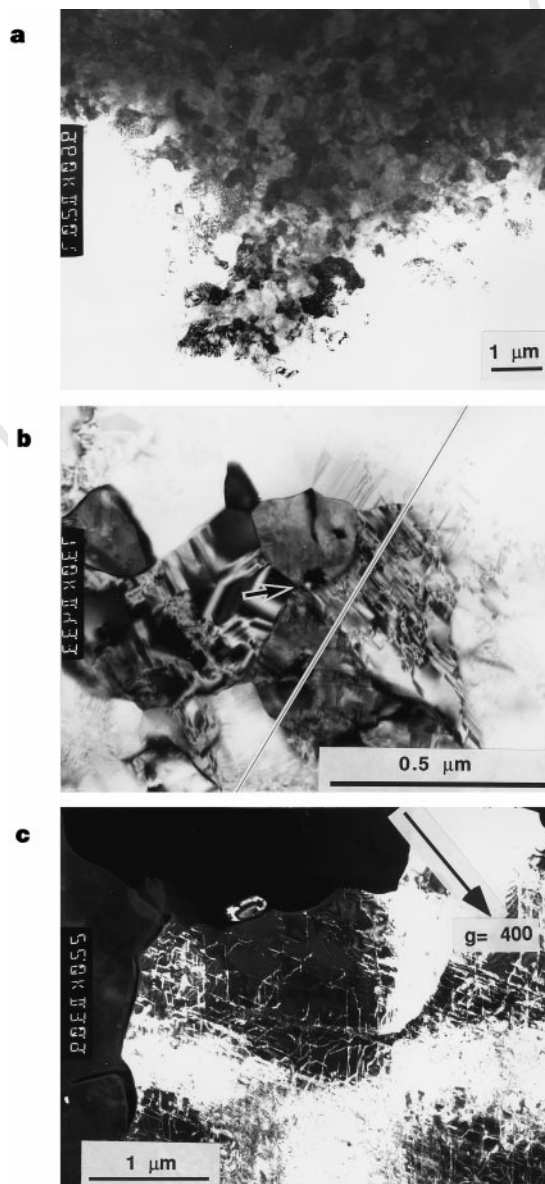


Figure 2 Transmission electron micrographs of deformed ringwoodite. **a**, Bright-field image of the fine-grained specimen no. 704 ($\sim 0.4 \mu\text{m}$), transformed and deformed at 16 GPa and 1,400 K, showing numerous stacking faults with a displacement vector $1/4\langle 110 \rangle$ but few dislocations. **b**, Bright-field image of sample no. 704, showing a four-grain junction (shown by an arrow) indicative of grain-switching events²⁸. A thin line indicates the trace of shear plane. **c**, Dark-field image (with diffraction vector $\mathbf{g} = 040$) of a relatively coarse-grained specimen no. 1740 ($\sim 5 \mu\text{m}$), transformed and deformed at 16 GPa, 1,600 K, showing numerous dislocations with $\mathbf{b} = 1/2\langle 110 \rangle$. The dislocation density is $\sim 10^{14} \text{ m}^{-2}$, which is consistent with a stress of ~ 1 GPa. The dominant slip system is inferred to be $1/2\langle 110 \rangle\{111\}$ with a subsidiary system $1/2\langle 110 \rangle\{100\}$ which are common to other spinels³⁰. These microstructural observations suggest that deformation in no. 1740 occurred mostly by dislocation creep whereas deformation in no. 704 was by superplasticity or diffusion creep (grain-boundary sliding accommodated either by diffusional mass transport or by dislocation glide/climb).

on small grains show little evidence of dislocations (Fig. 2a) although numerous {110} stacking faults are observed. These stacking faults are characteristic of the spinel structure^{9,10,27}. (2) A high density of dislocations ($\sim 10^{13}$ – 10^{14} m⁻²) is observed in large grains. (3) Grains show much smaller elongation than expected from the bulk sample strain (the aspect ratio of grains is ~ 1.1 compared to the expected value of 1.7 from sample strain). (4) Four-grain junctions are frequently observed (Fig. 2b). The orientations of the grain boundaries at these junctions are nearly parallel to the maximum shear directions (>50% of boundaries are within 15° of the maximum shear direction). Such four-grain junctions are expected to occur during grain switching events involving grain-boundary sliding which is characteristic of “superplasticity”²⁸. The observations (1), (3) and (4) suggest that much of the sample strain is accommodated by grain-boundary sliding, and hence a grain-size-sensitive creep mechanism (diffusion creep or “superplasticity”) is likely to have operated in this sample. The observation (2) indicates, however, that deformation of grains larger than ~ 3 μ m is largely accommodated by dislocation creep.

Another sample (no. 1740) was transformed to spinel and deformed at 1,600 K and 16 GPa for 10 h. The average grain size is 5 μ m and a high density of dislocations ($\sim 10^{14}$ m⁻²) is observed. Grains are much more elongated than those in sample no. 704 and have an average aspect ratio of ~ 1.8 ; the grain boundaries have

wavy morphology. Most of the dislocations are straight along {110} while others are straight along {100}. These dislocations have Burgers vectors $\mathbf{b} = (1/2)\langle 110 \rangle$ and are thus either screw, 60° dislocations, edge or 45° dislocations (Fig. 2c). Thus deformation occurred mostly by dislocation glide which predominantly involved the slip system $1/2\langle 110 \rangle\{110\}$, but $1/2\langle 110 \rangle\{100\}$ was also noted. The $1/2\langle 110 \rangle\{111\}$ slip system is commonly seen in nearly stoichiometric oxide and germanate spinels^{9,10}, and the $1/2\langle 110 \rangle\{100\}$ slip system is common to an early stage of creep in nearly stoichiometric alumina spinel deformed at high temperatures⁹. Similar microstructures (wavy grain-boundaries and a large aspect ratio) were observed in another sample (no. 710) that was deformed at similar conditions (grain size ~ 6 μ m, $T = 1,600$ K, $P = 16$ GPa for 5 h).

The deformation mechanisms in silicate spinel in the dislocation creep regime are very similar to those which occur in oxide and germanate spinels (Mg_2GeO_4 and Ni_2GeO_4)¹⁰ thus indicating that germanate spinels are good analogues of ringwoodite for dislocation creep. The creep strength of germanate spinels in the dislocation creep regime is about three times higher than that of olivine compared at the same strain rate^{3,10,11}. Therefore we conclude that the creep strength of a spinel-rich layer is significantly higher than that of an olivine-rich layer if both of them deform by dislocation creep^{3,10}.

We compare our deformation conditions with a deformation-mechanism map for spinel constructed from the existing data on germanate spinels (Mg_2GeO_4 and Ni_2GeO_4)^{10–12} (Fig. 3). The deformation conditions for sample no. 704 (small grains, ~ 0.4 μ m) are dominantly in the diffusion creep or superplasticity regime, and those for nos 710 and 1740 are in the dislocation creep regime as predicted from the data on germanate spinels. A few large grains in sample no 704 (~ 3 μ m), however, showed evidence of dislocation creep. We note that large and small grains in no. 704 were deformed at the same temperature and pressure, and yet microstructures suggest the operation of different mechanisms. Therefore the grain size at which the transition in the deformation mechanism occurs at these conditions must be between ~ 0.4 and ~ 3 μ m, which is consistent with other data.

We conclude that we have documented the first, to our knowledge, direct evidence for a change in deformation mechanisms in ringwoodite (silicate spinel) caused by changes in grain size and temperature under the pressure and temperature conditions of the transition zone of the Earth's mantle. The results are consistent with the predictions based on germanate analogues. Our results provide strong support for the idea that a large reduction in grain size associated with the olivine-to-spinel transformation in cold slabs would lead to significant rheological weakening^{13,14}. In contrast, typical hot regions of the transition zone where the grain size of spinel is large are likely to have significantly higher viscosities than the olivine-rich upper mantle.

The present study demonstrates that direct experimental studies of plastic deformation of deep-mantle minerals are now possible. This technique, however, has some limitations including the fact that differential stresses in these experiments are higher than those expected in the Earth's interior. Estimates of differential stress are indirect²⁴, and the rheological flow laws are still not well constrained. When these points are improved, this technique should provide important constraints on rheological properties and microstructural development under deep-Earth conditions, and should contribute to a better understanding of the dynamics of the Earth's deep interior. □

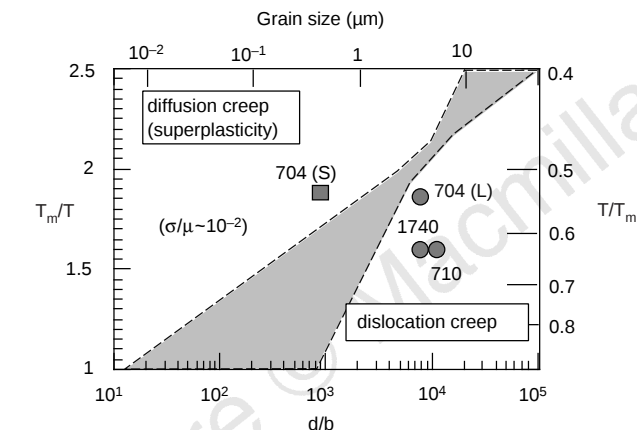


Figure 3 Deformation conditions of the present study as compared with deformation mechanism boundaries constructed from the data for germanate spinels (Mg_2GeO_4 and Ni_2GeO_4)^{10–12}. T , temperature; T_m , melting temperature; d , grain size; b , length of the Burgers vector; σ , stress; μ , shear modulus. Numbers next to symbols indicate run numbers (Table 1). For no. 704, both small grains (~ 0.4 μ m) and large grains (~ 3 μ m) are shown by 704 (S) and 704 (L) respectively. At relatively high temperatures and/or coarse grain size, dislocation creep dominates, whereas at relatively low temperatures and/or small grain size, diffusion creep or superplasticity dominates. A shaded area indicates the boundary between two mechanisms with uncertainties. A shaded square indicates a datum showing evidence for diffusion creep (superplasticity). Shaded circles indicate data showing evidence for dislocation creep (for details, see text). Scaling of rheological laws by homologous temperature (T/T_m) is assumed when comparing the data for germanates with those of $(\text{Mg},\text{Fe})_2\text{SiO}_4$ with $\text{Fe}/(\text{Fe} + \text{Mg}) = 0.4$. The transition grain size depends on stress as $d_{\text{ex}}/d_{\text{E}} = (\sigma_{\text{ex}}/\sigma_{\text{E}})^{1-n/m}$ where $d_{\text{ex}}, d_{\text{E}}$ are the transition grain size (stress level) in the experiments and in the Earth respectively, n is stress exponent ($n \approx 3$) and m is grain-size exponent ($m = 2-3$). In cold slabs ($T/T_m \approx 0.4$ and $\sigma_{\text{E}} \approx 100$ MPa), one gets $d_{\text{E}} = 0.5-2.0$ mm. Because theoretical modelling suggests a grain size of $\sim 1-10$ μ m in cold slabs¹⁴, grain-size-sensitive creep is likely to occur in cold slabs. In contrast, in hot regions away from slabs, where grain size is likely to be a few millimetres, either dislocation creep or grain-size-sensitive creep (such as diffusion creep) could occur²⁹. This implies that cold subducting slabs would lose their strength after the olivine-spinel transformation that occurs in the deep transition zone¹⁴, whereas the deep spinel-rich transition zone away from slabs could be stronger than the olivine-rich upper mantle.

Received 5 January; accepted 1 June 1998.

1. Forte, A. M. & Woodward, R. L. in *Earth's Deep Interior* (ed. Crossley, D. J.) 337–404 (Overseas Publishers Assoc./Gordon & Breach, Amsterdam, 1997).
2. Bunge, H.-P., Richards, M. A. & Baumgardner, J. R. The effect of viscosity stratification on mantle convection. *Nature* **379**, 436–438 (1996).
3. Karato, S. Plasticity-crystal structure systematics in dense oxides and its implications for the creep strength of the Earth's deep interior: A preliminary result. *Phys. Earth Planet. Inter.* **55**, 234–240 (1989).

4. Borch, R. S. & Green, H. W. II Dependence of creep in olivine on homologous temperature and its implication for flow in the mantle. *Nature* **330**, 345–348 (1987).
5. Peltier, W. R. in *Mantle Convection* (ed. Peltier, W. R.) 389–478 (Gordon & Breach, New York, 1989).
6. Mitrova, J. X. & Forte, A. M. Radial profile of mantle viscosity: Results from the joint inversion of convection and postglacial rebound observables. *J. Geophys. Res.* **102**, 2751–2769 (1997).
7. Kido, M. & Cadek, O. Inferences of viscosity from the oceanic geoid: Indication of a low viscosity zone below the 660-km discontinuity. *Earth Planet. Sci. Lett.* **151**, 125–137 (1997).
8. Ringwood, A. E. Phase transformations and their bearing on the constitution and dynamics of the mantle. *Geochim. Cosmochim. Acta* **55**, 2083–2110 (1991).
9. Duclos, R., Doukhan, N. & Escaig, B. High temperature creep behaviour of nearly stoichiometric alumina spinel. *J. Mater. Sci.* **13**, 1740–1748 (1978).
10. Dupas-Bruzek, C., Green, H. W. II, Doukhan, N., Doukhan, J.-C. & Tingle, T. N. The rheology of olivine and spinel magnesium germanate (Mg_2GeO_4): TEM study of the defect microstructures. *Phys. Chem. Miner.* (in the press).
11. Lawlis, J. R., Zhao, Y. & Karato, S. High temperature creep of nickel germanate spinel. *Eos* **79**, S331 (1998).
12. Vaughan, P. J. & Coe, R. S. Creep mechanism in Mg_2GeO_4 : effects of a phase transition. *J. Geophys. Res.* **86**, 389–404 (1981).
13. Rubie, D. C. The olivine $\rightarrow$ spinel transformation and the rheology of subducting lithosphere. *Nature* **308**, 505–508 (1984).
14. Riedel, M. R. & Karato, S. Grain-size evolution in subducted oceanic lithosphere associated with the olivine-spinel transformation and its effects on rheology. *Earth Planet. Sci. Lett.* **148**, 27–44 (1997).
15. van der Hilst, R. D., Engdahl, E. R., Spakman, W. & Nolet, G. Evidence for deep mantle circulation from global tomography. *Nature* **353**, 37–43 (1991).
16. Fukao, Y., Obayashi, M., Inoue, H. & Nishii, M. Subducting slabs stagnant in the mantle transition zone. *J. Geophys. Res.* **97**, 4809–4822 (1992).
17. Zhou, H. A high-resolution P wave model for the top 1200 km of the mantle. *J. Geophys. Res.* **101**, 27791–27810 (1996).
18. Machetel, P. & Weber, P. Intermittently layered convection in a model mantle with an endothermic phase-change at 670 km. *Nature* **350**, 55–57 (1991).
19. Peltier, W. R. & Solheim, L. P. Mantle phase-transitions and layered chaotic convection. *Geophys. Res. Lett.* **19**, 321–324 (1992).
20. Honda, S., Yuen, D. A., Balachandrar, S. & Reuteler, D. 3-dimensional instabilities of mantle convection with multiple phase-transitions. *Science* **259**, 1308–1311 (1993).
21. Tackley, P. J., Stevenson, D. J., Glatzmaier, G. A. & Schubert, G. Effects of an endothermic phase-transition at 670 km depth in a spherical model of mantle convection in the Earth's mantle. *Nature* **361**, 699–704 (1993).
22. Dupas, C., Doukhan, N., Doukhan, J.-C., Green, H. W. II & Young, T. E. Analytical electron microscopy of a synthetic peridotite experimentally deformed in the β olivine stability field. *J. Geophys. Res.* **99**, 15821–15832 (1994).
23. Dupas-Bruzek, C., Sharp, T. G., Rubie, D. C. & Durham, W. B. Mechanisms of transformation and deformation in $\text{Mg}_2\text{Fe}_{0.5}\text{SiO}_4$ olivine and wadsleyite under non-hydrostatic stress. *Phys. Earth Planet. Inter.* **108**, 33–48 (1998).
24. Karato, S. & Rubie, D. C. Toward an experimental study of deep mantle rheology: A new multi-anvil sample assembly for deformation studies under high pressures and temperatures. *J. Geophys. Res.* **102**, 20111–20122 (1997).
25. Ohtani, E. & Kumazawa, M. Melting of forsterite Mg_2SiO_4 up to 15 GPa. *Phys. Earth Planet. Inter.* **27**, 32–38 (1981).
26. Katsura, T. & Ito, E. The system $\text{Mg}_2\text{SiO}_4\text{--Fe}_2\text{SiO}_4$ at high pressure and temperature: precise determination of stabilities of olivine, modified spinel and spinel. *J. Geophys. Res.* **94**, 15663–15670 (1989).
27. Vaughan, P. J. & Kohlstedt, D. L. Cation stacking faults in magnesium germanate spinel. *Phys. Chem. Mineral.* **7**, 241–245 (1981).
28. Ashby, M. F. & Verrall, R. A. Diffusion accommodated flow and superplasticity. *Acta Metall.* **21**, 149–163 (1973).
29. Karato, S. in *Earth's Deep Interior* (ed. Crossley, D.) 223–272 (Overseas Publishers Assoc./Gordon & Breach, Amsterdam, 1997).

Acknowledgements. We thank H. Jung, Z. Wang, H. Schulze, H. Küfner and K.-H. Lee for technical assistance. This research was supported by grants from NSF, Alexander von Humboldt Stiftung, NATO and the EU "Human Capital and Mobility-Access to Large Scale Facilities" Programme.

Correspondence and requests for materials should be addressed to S.K. (e-mail: karato@maroon.tc.umn.edu).

Dynamic topography, plate driving forces and the African superswell

Carolina Lithgow-Bertelloni*† & Paul G. Silver*

* Department of Terrestrial Magnetism, Carnegie Institute of Washington, Washington DC 20015, USA

Discovering the connection between processes observed to occur at the surface of the Earth and its internal dynamics remains an essential goal in the Earth sciences. Deep mantle structure, as inferred from seismic tomography or subduction history, has been shown to account well for the observed surface gravity field and motions of tectonic plates^{1–3}. But the origin of certain

large-scale features, such as the anomalous elevation of the southern and eastern African plateaux, has remained controversial. Whereas the average elevation of most cratons is between 400 and 500 m, the southern African plateau stands more than 1 km above sea level, with the surrounding oceans possessing a residual bathymetry in excess of 500 m (ref. 4). Global seismic tomography studies have persistently indicated the existence of a large-scale low-velocity anomaly beneath the African plate^{5–10} and here we show that mantle flow induced by the density variations inferred from these velocity anomalies can dynamically support the excess elevation of the African 'superswell'. We also find that this upwelling mantle flow—which is most intense near the core-mantle boundary—constitutes a significant driving force for tectonic plates in the region.

Large-scale anomalous elevation of a continent can originate from: (1) changes in the average density and/or thickness of the lithosphere; and (2) vertical motion of the continent in the absence of faulting or folding. A significant source of vertical motion is dynamic topography¹¹.

Most previous explanations for the high residual topography of the southern and eastern African plateaux and surrounding oceans (Fig. 1a) have invoked isostatic mechanisms. These have included heating of the lithosphere, as indicated by high heat-flow measurements in the southern African craton¹², and lithospheric thinning, as evidenced by studies of the gravity field and Cenozoic volcanism under the eastern African plateau¹³. Although certainly some of these mechanisms will be at play, they have failed to successfully account for the existence of the African superswell.

Here we focus on dynamic topography as an explanation for the African superswell. Dynamic topography is a deformation of the surface of the Earth, supported by the vertical stresses at the base of the lithosphere, that are generated by flow in the mantle below^{1,14–19}. This is in contrast to the more familiar mechanism of isostatically supported topography which is in equilibrium at the Earth's surface and would exist even in a static mantle. Dynamic topography is expected to contribute to the elevation of the African continent because of upwelling (associated with anomalously low seismic velocities) that coincides with a high in the long-wavelength geoid¹. But we note that detection of dynamic topography, particularly over the oceans, remains controversial.

We predict the dynamic topography from a calculation of instantaneous flow in the Earth's mantle²⁰. The flow is a solution to the equations for the conservation of mass and momentum in an incompressible newtonian viscous fluid. The buoyancy forces that drive the flow are prescribed by a model of lateral density heterogeneity within the mantle. These calculations yield the radial stress acting at the base of the lithosphere, and allow us to compute the resulting dynamic topography simply by dividing the stress by the density contrast between the lithosphere and the overlying medium and by the gravitational acceleration. Because we are focusing on the dynamic topography of Africa, we use a density contrast ($3,200 \text{ kg m}^{-3}$) appropriate for the continent–air interface. This leads us to underestimate by $\sim 25\%$ the dynamic topography over ocean basins, for which the appropriate density jump is smaller because of the larger density of water.

We explore two different models of mantle density heterogeneity. The first is based on the history of subduction for the past 200 Myr. This type of model is based on the idea that in an Earth primarily heated from within, cold subducted slabs will be the main source of thermal buoyancy in the mantle and will dominate the structure of the flow. This model has been extremely successful at explaining the geoid², present and past plate motions³, and the history of uplift and subsidence of continents^{11,21}. In this model, upwellings are passive, and are characterized by the broad-scale return flow away from downwelling regions.

Active upwellings, generated in the basal thermal boundary layer, may also be an integral part of the flow. To address this possibility,

† Present address: Department of Geological Sciences, University of Michigan, Ann Arbor, Michigan 48109, USA.

we consider a second model of mantle density heterogeneity, one that is inferred from seismic tomography. This model will include a signal from active upwellings in addition to the slab signature. Global seismic tomography models have consistently imaged a coherent low-velocity anomaly under the African plate^{5–10}. One of the most recent high-resolution seismic models⁹ shows a continuous feature from the surface to the core–mantle boundary. In the lower mantle, between 1,000 and 2,890 km depth, the anomaly is located under southern Africa. In the shallower portions of the mantle, at depths less than 1,000 km, the anomaly becomes thinner and jogs northeastwards, under the Great Rift Valley (eastern African plateau). This type of thinning is expected for an upwelling encountering a lower-viscosity layer²²—the upper mantle has an

average viscosity more than an order of magnitude less than that of the lower mantle^{2,23–25}.

The density heterogeneity model based on subduction history does not adequately reproduce the residual topography (Fig. 1b). Instead, it produces a very broad-scale topographic high in the entire Atlantic basin, which is the result of the passive, upwelling return flow that balances the downwelling flow induced by subduction-related buoyancy. This broad upwelling cannot explain the anomalous elevation of the African superswell. If the anomalous elevation is to be related to dynamic topography, other density anomalies must be present in the mantle, in particular a concentrated low-density anomaly that we can associate with a large-scale upwelling.

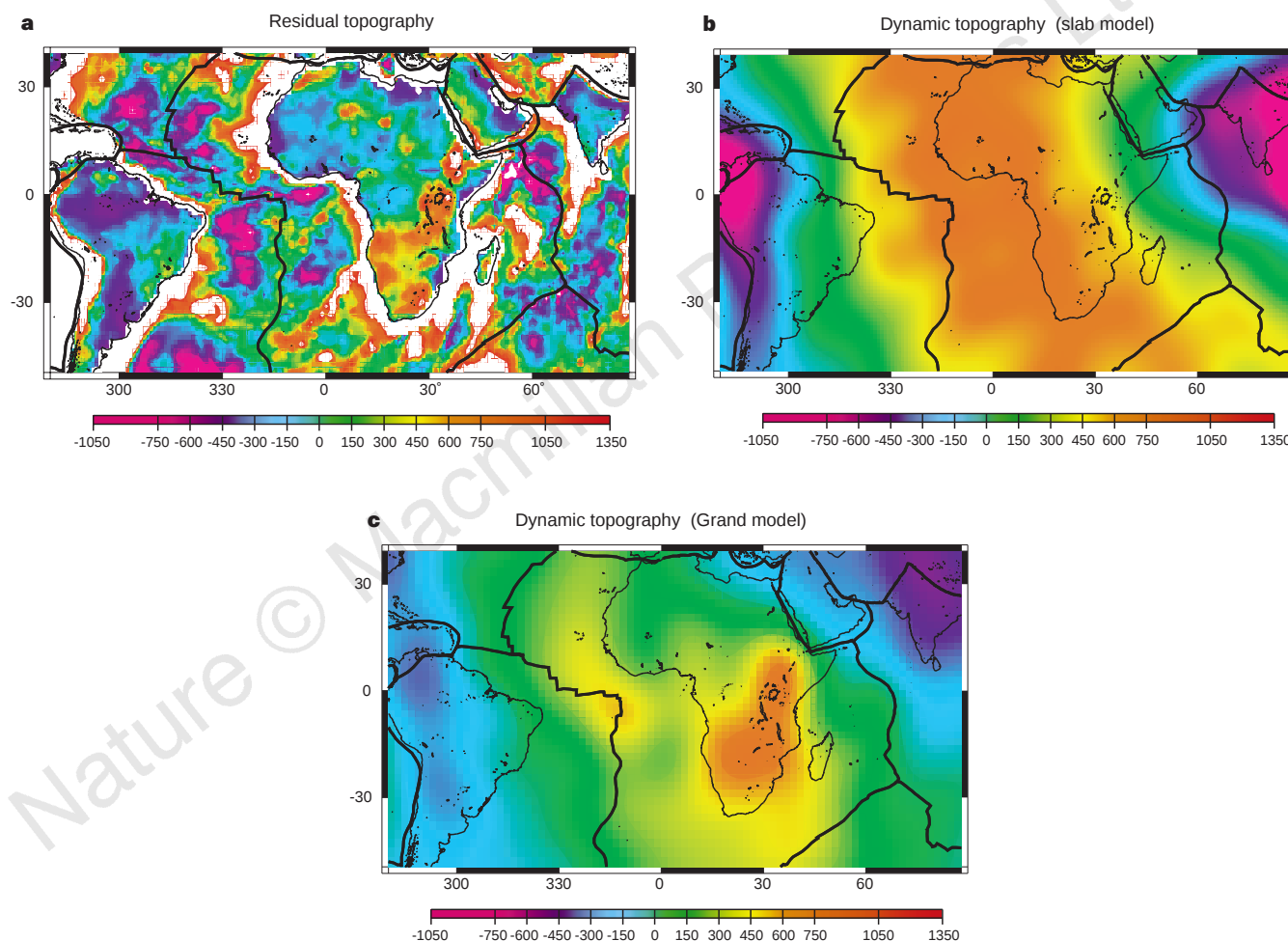


Figure 1 Comparison of the residual topography over Africa and the predicted dynamic topography given by the subduction history model and Grand's tomographic model⁹. **a**, Residual topography (in m) over the African continent and surrounding oceans. The residual bathymetry was obtained by subtracting a cooling half-space model from the observed values. The continental residual topography was obtained by subtracting a mean value of 565 m from the observed elevations⁴. Continental shelves are shown in white as are areas with residual topography higher than 1,350 m, whose anomalous topography or bathymetry is probably related to lithospheric effects. The reddish region and its extension to the surrounding oceans is what is known as the African superswell. Areas with excess topography over 1,000 m are much smaller in scale. **b**, Predicted dynamic topography (in m) over the African continent using the density heterogeneity from subduction history². Dynamic topography is independent of the absolute value of mantle viscosity, but is sensitive to the relative viscosity structure of the mantle. We choose a viscosity structure for the mantle where the lower mantle is 50 times more viscous than the upper mantle. This is

the viscosity structure found to best fit the geoid for this density heterogeneity model and one consistent with previous work^{2,25}. The surface density contrast is $3,200 \text{ kg m}^{-3}$. We note that the predicted topography does not match the observed residual topography over southern and eastern Africa in **a**, **c**, Predicted dynamic topography (in m) over the African continent using the density heterogeneity from seismic tomography⁹. We assume that density anomalies are thermal in nature and can be simply related to seismic velocities. We convert from velocity to density using a constant factor of $0.4 \text{ kg m}^{-3} \text{ km}^{-1} \text{ s}$. We note the excellent agreement between the predicted pattern of dynamic topography over the southern and eastern African plateaux and the observed residual topography. The global peak-to-peak amplitude is 1,350 m, and the r.m.s. amplitude is 230 m. Density contrast and viscosity structure as for **b**. Shown are harmonic degrees 1–15; power at higher harmonics is insignificant ($<20\%$ of total). The geoid predicted by Grand's model yields a variance reduction of 55% for our viscosity structure, comparable to other tomographic models.

To calculate the dynamic topography induced by the tomographically inferred density field, we use only the velocity anomalies below 325 km depth. At depths less than 325 km many of the anomalies found in the upper mantle are associated with cratons, and are probably chemical rather than thermal in nature. To avoid any contamination of our results with deep lithospheric structure, we restrict ourselves to depths greater than 325 km. We convert the remaining velocity anomalies to density by using a constant conversion factor, $0.4 \text{ kg m}^{-3} \text{ km}^{-1} \text{ s}$, from 325 km to the core–mantle boundary. This is a value that is well within the bounds of laboratory experiments and geodynamical models²⁶. We find that the dynamic topography predicted from this density field reproduces remarkably well the anomalous residual topography over the African plate (Fig. 1c). The anomalous topography under the eastern African plateau is the result of the portion of the density anomaly in the upper mantle (325–525 km). The topographic high corresponding to the southern African plateau is the direct result of the density anomaly at depths greater than 1,000 km. Without including the lower-mantle anomaly (below 1,000 km), the pattern and amplitude of dynamic topography is not reproduced.

Our results agree in amplitude, as well as pattern, with the observed residual topography (peak predicted amplitudes $\sim 700 \text{ m}$, average residual elevation of the African craton $\sim 650 \text{ m}$). The agreement in amplitude may be partly fortuitous, because there are two known sources of amplitude bias that will tend to counteract each other. First, intentionally conservative seismic anomalies in Grand's model⁹ will underpredict the amplitude of velocity heterogeneity in the mantle and consequently the amplitude of surface dynamic topography. Dynamic topography calculated with this model in other regions (such as the circum-Pacific subduction zones) is smaller than that predicted by other models^{2,27} by as much as 650 m. Second, mantle flow calculations tend to overpredict dynamic topography. Resolving the discrepancy between models of dynamic topography and its observed amplitude remains a fundamental problem in geodynamics. One recent approach has been to allow the development of dynamic topography at internal phase or chemical boundaries^{17–19}.

In this context it is important to note that the geoid predicted by Grand's model agrees as well with observations as most other tomographic models for the same parametrization and boundary conditions. The agreement is good both at the global and regional

level, and the model correctly predicts the geoid high over Africa (see Supplementary Information).

With the caveat that mantle flow models tend to overpredict the magnitude of dynamic topography, we suggest that the deep low-velocity anomaly under Africa, as seen consistently in tomographic models, is the primary cause of the African superswell. The shallower portions of this lower-mantle feature are at least partly responsible for the high elevations of the eastern African plateau. This deep-mantle feature and its upper-mantle extension might also be the source for the high heat flow and volcanism in this region.

Having established the effect that mantle flow has on surface topography, we turn our attention to the possible contribution of a large-scale upwelling to plate driving forces. Such forces are, in general, gravitational in nature and result from the large lateral density contrasts that arise from the cooling of the oceanic lithosphere and its subsequent subduction. But, because plates are coupled to mantle flow, other density anomalies in the mantle (such as the large-scale upwelling described here) will also have an influence on the shear tractions at the base of the lithosphere. To consider its effect, we separate plate driving forces into four categories; lithospheric thickening, upper-mantle slabs, lower-mantle slabs, and a general large-scale upwelling corresponding to the low velocity anomalies in Grand's tomographic model. We calculate the instantaneous flow from each driving load (that is, density distribution) independently, compute the shear tractions acting at the base of the lithosphere and find, in the no-net-torque approximation, the driving torques exerted on each plate³.

For plates in the Atlantic basin we find that the torque magnitudes related to the large-scale upwelling are large when compared to other driving forces—in fact, much larger (almost an order of magnitude) than lithospheric thickening and comparable to the torques from upper-mantle slabs (Fig. 2). Most of the large-scale upwelling contribution ($\sim 90\%$) is derived from the lower-mantle portion of the anomaly. Clearly then, the driving torques exerted by a large-scale upwelling, even one constrained to the lower mantle, cannot be ignored, provided that there is no barrier to flow between upper and lower mantle. Given that there is a significant and visible contribution to the radial stresses from this upwelling, this result is not unexpected.

We have shown here that the low-velocity anomaly under the African plate, seen in tomographic models, can be viewed as an image of a large-scale active upwelling under this region. This density anomaly induces a surface boundary deformation (dynamic topography) that explains the existence of the African superswell including the southern African plateau and surrounding oceans, and the eastern African plateau. The upwelling also constitutes a significant driving force for plates in its vicinity. The upwelling is most intense in the vicinity of the core–mantle boundary, presumably its point of origin. We therefore conclude that large-scale upwellings originated from the core–mantle boundary, such as the one we have examined here, are important features of global mantle flow, and have a strong influence on surface processes. Such upwellings were probably also important in the past. As previously shown^{11,21}, the inferred evolution of dynamic topography predicted from known sources of buoyancy, such as those related to the history of subduction, can explain much of the uplift history implied by the Cenozoic continental flooding record^{11,21}. It may now be possible to similarly examine the evolution of deep-mantle upwellings by comparison with the uplift history of the African plate, in a manner similar to that suggested by White and Lovell for plume activity²⁸.

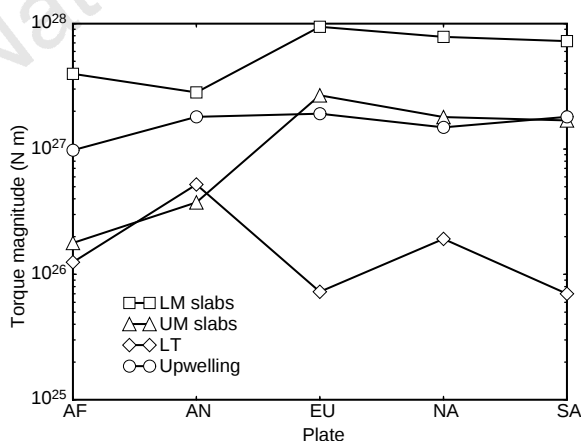


Figure 2 Torque magnitudes of individual driving forces for plates in the Atlantic basin. Plate abbreviations as follows: AF, Africa; AN, Antarctica; EU, Eurasia; NA, North America; SA, South America. Every calculation is referenced to the same absolute upper-mantle viscosity (10^{21} Pa s). We note that the torque magnitudes for the large-scale upwelling ('Upwelling') are very large, an order of magnitude larger than lithospheric thickening ('LT') and comparable to upper mantle slabs ('UM slabs'). The contribution of lower-mantle slabs ('LM slabs') is shown for reference (open squares).

Received 30 March; accepted 25 June 1998.

- Hager, B. H. *et al.* Lower mantle heterogeneity, dynamic topography and the geoid. *Nature* **313**, 541–554 (1985).
- Ricard, Y., Richards, M., Lithgow-Bertelloni, C. & LeStunff, Y. A geodynamical model of mantle density heterogeneity. *J. Geophys. Res.* **98**, 21895–21909 (1993).
- Lithgow-Bertelloni, C. & Richards, M. A. Cenozoic plate driving forces. *Geophys. Res. Lett.* **22**, 1317–1320 (1995).

Correspondence and requests for materials should be addressed to C.L.-B. (e-mail: crlb@umich.edu).

Episodic-like memory during cache recovery by scrub jays

Nicola S. Clayton* & Anthony Dickinson†

* *Section of Neurobiology, Physiology & Behavior, University of California at Davis, California 95616, USA*

[†] *Department of Experimental Psychology, University of Cambridge, Downing Street, Cambridge CB2 3EB, UK*

The recollection of past experiences allows us to recall what a particular event was, and where and when it occurred^{1,2}, a form of memory that is thought to be unique to humans³. It is known, however, that food-storing birds remember the spatial location⁴⁻⁶ and contents⁶⁻⁹ of their caches. Furthermore, food-storing animals adapt their caching and recovery strategies to the perishability of food stores¹⁰⁻¹³, which suggests that they are sensitive to temporal factors. Here we show that scrub jays (*Aphelocoma coerulescens*) remember 'when' food items are stored by allowing them to recover perishable 'wax worms' (wax-moth larvae) and non-perishable peanuts which they had previously cached in

visuospatially distinct sites. Jays searched preferentially for fresh wax worms, their favoured food, when allowed to recover them shortly after caching. However, they rapidly learned to avoid searching for worms after a longer interval during which the worms had decayed. The recovery preference of jays demonstrates memory of where and when particular food items were cached, thereby fulfilling the behavioural criteria for episodic-like memory in non-human animals.

Scrub jays in the Degrade group were given a series of pretraining trials (see Methods) in which they learned that worms degrade and become unpalatable over time; they therefore learned to avoid recovering these items when a relatively long time (124 h) had elapsed between caching and recovery. The birds then received a further pair of training trials in which they were required to cache peanuts in one side of a distinctive, sand-filled storage tray during one caching phase, and wax worms in the opposite side during the other caching phase (Fig. 1). Different, novel trays were used on every trial throughout the experiment so that the cache sites were trial-unique. The two caching phases were separated by 120 h, and birds were subsequently allowed to recover items from both sides of the caching tray 4 h after the second caching phase. The 4-h and 124-h trial designations refer to the length of the time that elapsed between caching and recovering the wax worms: on the 124-h trial, the birds first stored wax worms 124 h before cache recovery and then cached peanuts in the other side of the tray 4 h before recovery; whereas on the 4-h trial, birds cached the wax worms 4 h before recovery, having previously stored peanuts 124 h prior to recovery. As the wax worms were still fresh at recovery on the 4-h trial, and as the fresh worms are their preferred food, the birds directed more of their recovery inspections to the worm side of the storage tray on this trial. By contrast, most of their recovery inspections were directed to the peanut side of the tray on the 124-h trial when the worms were in a decayed state.

During both pretraining and training trials, birds could use the sight and smell of their caches as cues about where to search during cache recovery. To test whether or not jays could remember if it was the worms or the peanuts that had been cached, and where and when they had been cached, each bird received a pair of test trials that differed from the previous pair of training trials in that all food items were removed before the recovery phase of each test trial and fresh sand was placed in the tray. This procedure ensured that the birds had to rely on memory during cache recovery, because olfactory and visual cues about the contents of caches were no longer available to the bird. If jays can remember when and where they stored the two types of food, they should show a preference for

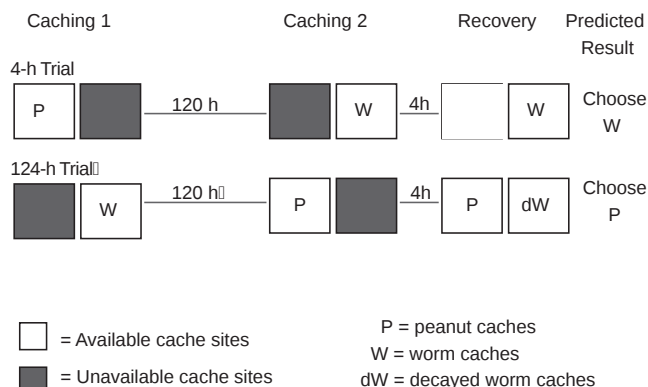


Figure 1 Conditions and contents of the caching tray on different phases of 4-h and 124-h training and trials. The trial designation refers to the length of the time that elapsed between caching and recovering the wax worms. During the caching phases, birds were prevented from storing food items in the shaded halves of the tray by a cover, but they were free to cache in the open, non-shaded halves.

sites at which the worms had recently been cached on the 4-h trial because they should expect the favoured wax worms still to be fresh. We assessed this preference by recording the sites that the birds inspected by probing the sand substrate with their bills. This preference should be reversed on the 124-h trial if the birds can also recall that the wax worms had been cached a relatively long time ago and therefore would have decayed and become unpalatable.

The results of the test trials fulfilled both predictions. Eighty per cent of birds in the Degrade group directed their first inspection to

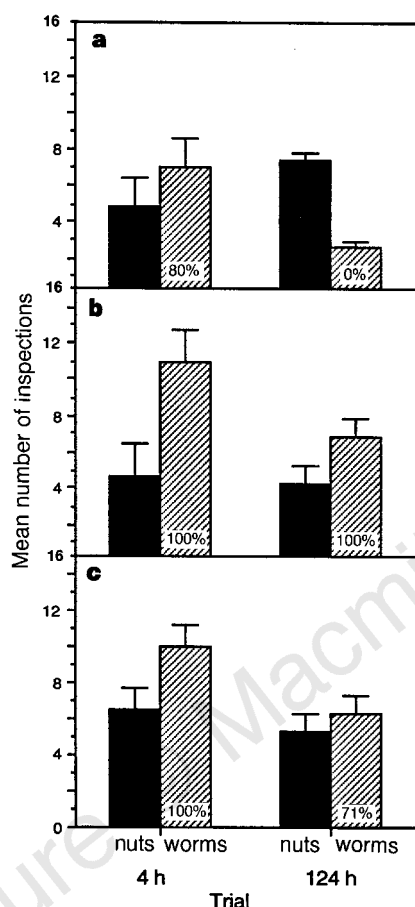


Figure 2 Mean number of inspections directed to the peanut and worm sides of the storage trays (error bars represent mean standard error for these contrasts) during the recovery phase of 4-h and 124-h test trials. All birds inspected the worm side more than the peanut side on the 4-h trial ($F_{1,14} = 9.76, P < 0.01$), a preference that does not differ between the groups ($F_{2,14} = 0.90, P > 0.25$). By contrast, there is a significant interaction between group and side for the number of inspections on the 124-h trial ($F_{2,14} = 9.59, P < 0.01$), and the Degrade group (a) inspected the peanut side more than worm side ($F_{1,14} = 12.79, P < 0.01$), whereas the Replenish group (b) inspected the worm side more than the peanut side ($F_{1,14} = 4.74, P < 0.05$). Birds in the Pilfer group (c) showed no significant preference ($F_{1,14} = 0.66, P > 0.25$). The groups also differ in the number of inspections to the worm side of the storage tray ($F_{2,14} = 6.26, P < 0.05$) but not to the peanut side ($F_{2,14} = 1.34, P > 0.25$). A post-hoc Newman-Keuls test reveals that the Degrade group made fewer inspections to the worm side than the other two groups ($P < 0.05$), which did not differ ($P > 0.05$). Also shown is the percentage of birds in each group that made their first inspections to the worm side of the storage tray. There is no difference between the groups in the distribution of first inspections on the 4-h trial ($\chi^2_2 = 2.55, P > 0.20$) and, overall, more first inspections were directed to the worm side than the peanut side (binomial test: $P < 0.01$). The groups differ, however, in their distribution of first inspections on the 124-h trial ($\chi^2_2 = 11.50, P < 0.01$), and one-tail binomial tests reveal that more first inspections were directed at the peanut side by the Degrade group and to the worm side by the Replenish group ($P < 0.05$ in both cases), whereas the Pilfer group showed no significant bias to either side of the caching tray ($P > 0.30$).

the worm side of the caching tray on the 4-h trial, whereas all birds inspected the peanut side first on the 124-h trial (Fig. 2a). A similar pattern was seen for the total number of inspections during cache recovery, with most inspections being directed to the worm side on the 4-h trial but to the peanut side on the 124-h trial (Fig. 2a). This cache recovery pattern suggests that the birds remembered not only where the worms were stored, but also information about the relative time between caching and recovery, in the sense that they searched preferentially in the worm side of the tray when the worms had been cached 4 h previously, but preferred to search in the peanut side when the worms have been cached 124 h before the recovery test.

To check that the preference for peanut caches on the 124-h trial was not due to more rapid forgetting of worm versus peanut caches, we compared the performance of birds in the Degrade group to that of a second, Replenish group. Birds in the Replenish group received the same training as those in the Degrade group except that the stored wax worms were removed immediately after the caching phase and replacing by fresh ones just before the cache recovery phase. Thus, birds in the Replenish group never had the opportunity to learn that worms decay over time. In contrast to the recovery pattern shown by the Degrade group (Fig. 2a), all birds in the Replenish group directed their first inspection to the worm side of the caching tray on both the 4-h and the 124-h test trial (Fig. 2b). Although the total number of inspections to the worm side was reduced on the 124-h trial relative to the 4-h trial, the Replenish group inspected the worm side more frequently than the peanut side during cache recovery on both trials (Fig. 2b), thereby demonstrating that the peanut-side preference shown by the Degrade group on the 124-h trial was not simply due to differential forgetting of worm caches.

We also investigated another ecologically inspired procedure for training the birds to encode and retrieve temporal information about caching. This procedure capitalized on the fact that, in the wild, increasing the time between storage and recovery enhances the likelihood that the caches will be pilfered (stolen) by another animal. In response to pilfering, birds adjust both the locations in which they search for caches and those in which they store food subsequently^{14–17}. To teach our birds that the likelihood that the worm caches would be pilfered increased with time since caching, we trained the Pilfer group on a similar procedure to the Degrade group, but in this case the wax worms were removed from the storage tray on the 124-h training trial so that the Pilfer group could recover only peanut caches. In common with birds in the Replenish and Degrade group, however, birds in the Pilfer group could recover both peanut and worm caches on the 4-h training trial.

During cache recovery on the 4-h test trial, birds in the Pilfer group showed the same preference for the worm sites as the other groups as they all made their first inspection to the worm side (Fig. 2c), a bias that was also expressed in the total number of inspections made to worm and peanut sides (Fig. 2c). The critical results, however, are those for the 124-h test trial. If scrub jays can learn that the probability that caches will be pilfered increases with time that has elapsed between caching and recovery by remembering relatively how long ago they stored the wax worms, then their preference for inspecting the worm sites should be reduced on the 124-h trial relative to that shown by the Replenish group. This prediction was upheld: in contrast to the Replenish group, 29% of the birds in the Pilfer group made their first inspection to the peanut side on the 124-h trial (Fig. 2c), and there was no significant difference in the total number of inspections directed to the two sides of the storage tray (Fig. 2c).

In conclusion, the switch in preference from the worm side on the 4-h trial to the peanut side on the 124-h trial shown by birds in the Degrade group, and to a lesser extent by the Pilfer group, can only be explained by recall of three types of information: (1) 'what' items (peanuts and worms) were cached; (2) 'where' each type of

item was stored (left or right sides); and (3) 'when' (4 h or 124 h ago) the worms were cached. Current theories of human episodic memory refer to autonoetic consciousness³—the conscious experience of self—that accompanies episodic recall but, as this state has no obvious manifestation in non-linguistic behaviour³ it is probably undetectable in many species. In terms of purely behavioural criteria, however, the cache recovery pattern of scrub jays fulfils the three, 'what', 'where' and 'when' criteria for episodic recall and thus provides, to our knowledge, the first conclusive behavioural evidence of episodic-like memory in animals other than humans. □

Methods

Subjects. Adult, hand-raised scrub jays, which were allocated randomly to the Degrade ($n = 8$), Replenish ($n = 8$) and Pilfer ($n = 7$) groups, were maintained under the same conditions as in the previous studies in which they had participated^{8,18}. Six birds (Degrade, $n = 3$; Replenish, $n = 1$; Pilfer, $n = 1$) that failed to store at least one item of both food types during the test trials were omitted from the analysis.

Training and testing. The birds were deprived of their maintenance diet for 4 h before each caching phase of the 4-h and 124-h trials (Fig. 1), and during the 4-h retention interval between the second caching phase and recovery. At the start of each caching phase, a storage tray was placed in the bird's home cage together with a bowl of 50 shelled peanuts or wax worms for 15 min. To provide a storage site that was unique and discriminable on every trial for each bird, the storage tray used on a particular trial was drawn from a large set of sand-filled plastic ice-cube trays (2×7 array of 2.5-cm cube molds), each rendered visuospatially distinct by a surrounding structure of Lego bricks⁸. Access to one side of the storage tray during the first caching phase and the other side during the second caching phase was prevented by a clear Perspex cover. At the end of each caching phase, the number and location of caches were recorded before the food items were removed from the tray. Before each 15-min recovery phase, peanuts were placed in the same ice-cube molds in which they had been cached on the training trials. On 4-h training trials, wax worms were also returned to their original cache locations. On 124-h training trials, decayed worms were placed in those sites for birds for the Replenish group. The worm sites remained empty for the Pilfered group. Furthermore, before all recovery phases the sand substrate was replaced to remove any local visual cues about the location of caches. The procedure was the same on the test trials, except that none of the caches was returned to the storage tray before the recovery phase. The order of the 4-h and 124-h training and test trials was counterbalanced across birds within each group, as was the side of the storage tray in which peanuts and worms were cached. To minimize observer bias, different observers recorded the behaviour during the caching and recovery phases so that the observer during the recovery phase was unaware of the treatment received during caching.

Pretraining. Before the training and test trials, all birds received at least four pairs of pretraining trials. The procedure during pretraining trials differed from that during training trials in three respects. (1) On pretraining trials, peanuts and worms were cached in separate, visuospatially distinct, trial-unique storage trays. Thus during cache recovery, birds were presented with a choice between the two trays on pretraining trials. (2) The second caching phase of a pretraining trial followed immediately after the first one. (3) The cache recovery phase occurred 4 h after the second caching phase on one trial of each pair of pretraining trials and 124 h later on the other pretraining trial. The order in which peanuts and worms were cached was counterbalanced within groups, as was the order of the 4-h and 124-h trials within each pair of pretraining trials.

Received 5 May; accepted 7 July 1998.

1. Tulving, E. in *Organization of Memory* (eds Tulving, E. & Donaldson, W.) 382–403 (Academic, New York, 1972).
2. Tulving, E. *Elements of Episodic Memory* (Oxford Univ. Press, New York, 1983).
3. Tulving, E. & Markowitsch, H. J. Episodic and declarative memory: role of the hippocampus. *Hippocampus* (in the press).
4. Shettleworth, S. J. in *Proceedings of NATO Advanced Study Institute Series Maratea, Italy* (eds Alleva, E., Fasolo, A., Lipp, H.-P. & Nadel, L.) 158–179 (Kluwer, The Hague, 1995).
5. Shettleworth, S. J. & Krebs, J. R. How marsh tits find their hoards: the roles of site preference and spatial memory. *J. Exp. Psychol. Anim. Behav. Proc.* **8**, 354–375 (1982).
6. Kamil, A. C. & Balda, R. P. Cache recovery and spatial memory in Clark's nutcrackers (*Nucifraga columbiana*). *J. Exp. Psychol. Anim. Behav. Proc.* **11**, 95–111 (1985).

7. Sherry, D. F. Food storage, memory, and marsh tits. *Anim. Behav.* **30**, 631–633 (1982).
8. Clayton, N. S. & Dickinson, A. Memory for the content of caches by scrub jays (*Apelocoma coerulescens*). *J. Exp. Psychol. Anim. Behav. Proc.* (in the press).
9. Kamil, A. C. & Balda, R. P. Differential memory for cache sites by Clark's nutcrackers (*Nucifraga columbiana*). *J. Exp. Psych. Anim. Behav. Proc.* **16**, 162–168 (1990).
10. Clayton, N. S., Mellor, R. & Jackson, A. Seasonal patterns of food storing in the European jay (*Garrulus glandarius*). *Ibis* **138**, 250–255 (1996).
11. Hadj-Chikh, L. Z., Steele, M. A. & Smallwood, P. D. Caching decisions by grey squirrels: a test of the handling time and perishability hypotheses. *Anim. Behav.* **52**, 941–948 (1996).
12. Reichman, O. J. Caching behaviour by eastern woodrats, *Neotoma floridana*, in relation to food perishability. *Anim. Behav.* **36**, 1525–1532 (1988).
13. Gendron, R. P. & Reichman, O. J. Food perishability and inventory management: a comparison of three caching strategies. *Am. Nat.* **145**, 948–968 (1995).
14. Stevens, T. A. & Krebs, J. R. Retrieval of stored seeds by marsh tits (*Parus palustris*) in the field. *Ibis* **128**, 513–515 (1984).
15. Hampton, R. R. & Sherry, D. F. The effects of cache loss on choice of cache sites in the black-capped chickadee. *Behav. Ecol.* **5**, 44–50 (1994).
16. Baker, M. C. & Anderson, P. Once-pilfered cache sites not avoided by black-capped chickadees. *Anim. Behav.* **49**, 1599–1602 (1995).
17. Kamil, A. C., Balda, R. P., Olson, D. J. & Good, S. Returns to emptied cache sites by Clark's nutcrackers, *Nucifraga columbiana*: a puzzle revisited. *Anim. Behav.* **45**, 241–252 (1993).
18. Clayton, N. S. & Dickinson, A. Motivational control of caching behaviour in the scrub jay *Aphelocoma coerulescens*. *Anim. Behav.* (in the press).

Acknowledgements. This work was supported by grants from the NIH and the Whitehall Foundation to N.S.C. and from the BBSRC to A.D. We thank A. Quagliano, M. Caun, J. Greig, D. Griffiths, K. Sanford, E. Krebs and G. Krebs for their help with aspects of the experiment; E. Tulving for discussion and inspiration; and K. Britten, R. Gerlai, J. Krebs, D. Lee, P. Marler, L. Miyasato, R. Morris, K. Sanford, E. Tulving and A. Quagliano for comments on the manuscript.

Correspondence and requests for materials should be addressed to N.S.C. (e-mail: nsclayton@ucdavis.edu).

Primary motor cortex is involved in bimanual coordination

O. Donchin, A. Gribova, O. Steinberg, H. Bergman & E. Vaadia

Department of Physiology and the Center for Neural Computation, The Hebrew University, Hadassah Medical School, Jerusalem 91120, Israel

Many voluntary movements involve coordination between the limbs^{1,2}. However, there have been very few attempts to study the neuronal mechanisms that mediate this coordination. Here we have studied the activity of cortical neurons while monkeys performed tasks that required coordination between the two arms. We found that most neurons in the primary motor cortex (MI) show activity specific to bimanual movements (bimanual-related activity), which is strikingly different from the activity of the same neurons during unimanual movements. Moreover, units in the supplementary motor area (SMA; the area of cortex most often associated with bimanual coordination³) showed no more bimanual-related activity than units in MI. Our results challenge the classic view that MI controls the contralateral (opposite) side of the body and that SMA is responsible for the coordination of the arms. Rather, our data suggest that both cortical areas share the control of bilateral coordination.

Common experience and psychophysical studies indicate that the limbs coordinate automatically, and that requirements altering this natural coordination increase task difficulty^{1,2,4}. Anatomical, behavioural and electrophysiological evidence indicate that SMA is an important contributor to bimanual coordination^{5–9}. However, there are reports of only transient disturbance in interlimb coordination with SMA lesions, suggesting that SMA is not solely responsible for bimanual coordination⁴.

We reinvestigated the roles of MI and SMA in bimanual coordination. Monkeys operated two X–Y manipulanda, one with each arm, which moved cursors on a video screen (see Methods and Fig. 1). In bimanual trials the two arms started to move together quite accurately, with an average inter-arm interval (IAI) of 16 ms (s.d. 74 ms) for monkey F and 21 ms (s.d. 56 ms) for monkey G, and

Table 1 Responsiveness and arm preference of MI and SMA cells in unimanual movements.

Area	Fraction of directionally selective cells		Contralateral versus ipsilateral preference of directionally selective cells				Total
	Total	Selective	Contra	Both (C > I)	Both (I > C)	Ipsi	
MI (F)	72	50 (69%)	26 (52%)	13 (26%)	1 (2%)	10 (20%)	50 (100%)
MI (G)	44	22 (50%)	10 (45%)	4 (18%)	3 (13%)	5 (22%)	22 (100%)
SMA (F)	73	43 (59%)	14 (32%)	8 (19%)	6 (14%)	15 (35%)	43 (100%)

Neurons that responded in relation to both ipsilateral and contralateral arm movements were divided into two groups: those showing stronger modulations in relation to the contralateral arm (C > I), and those showing stronger modulations in relation to the ipsilateral arm (I > C).

reached the targets with an average IAI of 5 and 15 ms (s.d., 106 and 125 ms). These IAIs are much shorter than required for successful performance, indicating synchronization of the movements of the hands.

We recorded single-unit activity from homologous sites in the two hemispheres during task performance. We analysed the activity of 189 neurons that were within the arm area of MI (72 from monkey F and 44 from monkey G) and SMA (73 from monkey F).

Most MI cells (62%, 72 of 116) were directionally selective (see Methods) during unimanual movements of either the contralateral or ipsilateral arm. Table 1 shows that, although many of these neurons (29%, 21 of 72) were activated in relation to movements of both arms, modulation in contralateral movements was stronger in 73% (53 of 72) of the responsive cells. In contrast, directionally selective units in the SMA (59% of the cells, 43 of 73) were distributed evenly between the two arms.

Cell activity revealed bimanual-related components of activity (Figs 2 and 3). Figure 2 shows the activity of a neuron recorded in the right SMA. It was inactive in unimanual movements towards 45° or 225° (Fig. 2, contralateral and ipsilateral columns), but it fired when the two arms moved in parallel towards 225° (Fig. 2b). It was less active in one other kind of bimanual trial (Fig. 2c) and failed to respond in the rest (Fig. 2a, d).

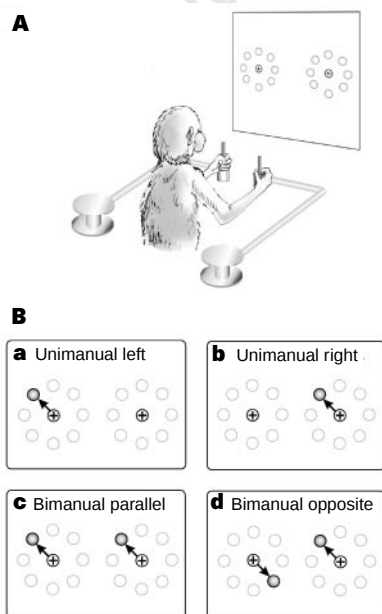


Figure 1 Behavioural paradigm and task. **A**, The monkey sits in a primate chair holding two manipulanda and facing a video screen. Two cursors (+) indicate the location of the manipulanda. Each cursor appears in the corresponding origin. All target locations are shown as grey circles surrounding each origin. **Ba–d**, Examples of the four main types of trials (the open circles are not visible to the monkey). In the examples shown the selected direction was 135°. The monkey performed these four types along with complementary four types of movement in the opposite direction (315°).

The neuron in Fig. 3A is from right MI, and fires before movements of the contralateral arm towards 45° (Fig. 3Aa, contralateral column). However, both arms together evoked a different pattern of activity (Fig. 3Aa, bimanual column). The late bimanual-related activity was almost completely absent when the monkey performed similar unimanual movements.

Many cells exhibited bimanual-related activity: 69% (53 of 72 in monkey F and 27 of 44 in monkey G) in MI and 64% (47 of 73) in SMA. This indicates that bimanual-related activity is at least as common in MI as in SMA. Note that these cells outnumber 'directionally selective' cells, suggesting that even classically 'unresponsive' neurons may participate in bimanual coordination. Table 2 shows the significance of many bimanual-related components, even by extremely strict criteria. The components in Figs 2

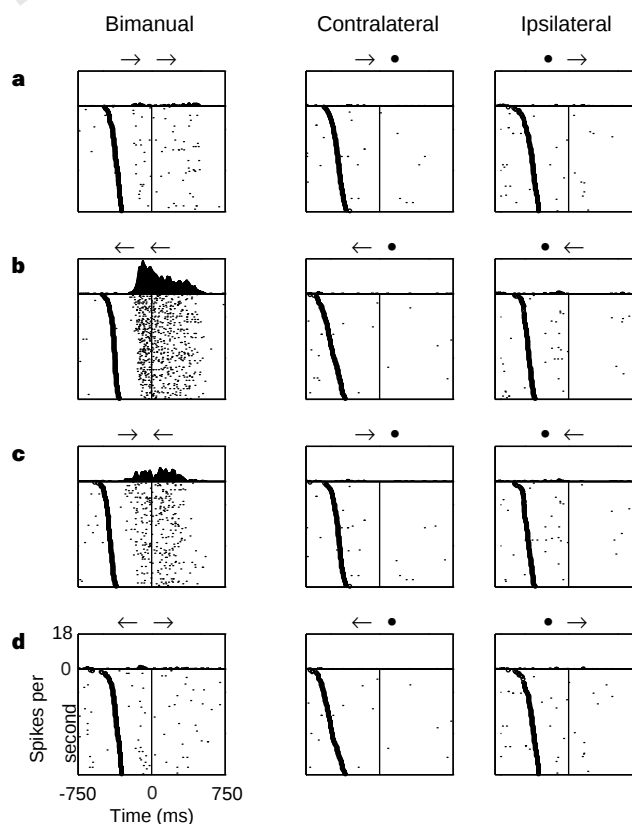


Figure 2 SMA cell with bimanual-related activity. **a–d**, Each row contains peristimulus time histograms (PSTHs) and raster displays depicting the cell activity in one type of trial. The direction of movement of each arm is indicated above the graph by arrows or a black circle if the arm does not move. The cell had strong activation only during bimanual movements (**b**, left column) and no activity in unimanual trials (right column is unimanual right; middle column is unimanual left). Trials are aligned on the beginning of the movement (of the first arm) and sorted by reaction time. The target onset is indicated by circles. The number of trials ($n = 127$) and the PSTH scale are identical in all plots. The movement directions were 45° (→) and 225° (←).

and 3A are significant at a level of $P < 10^{-16}$. The cell in Fig. 3B demonstrates a weak bimanual-related component at the limit of statistical significance (10^{-3}) recorded in the left MI. Examination of the strength (Table 2), latency and duration of bimanual-related components in the two areas failed to reveal substantive differences in the character of bimanual-related activity in MI and SMA.

Observation of the monkeys during task performance did not reveal postural adjustments that distinguished movements during bimanual and unimanual trials. Visual inspection of the recorded trajectories, velocity profiles and electromyogram (EMG) showed that bimanual and unimanual movements were similar in many trials, although there were differences. For example, an analysis of EMG patterns similar to the one performed on the neural activity revealed that 24% of the patterns showed statistically significant differences (as opposed to 69% of the neurons). The large number of trials in each condition allowed an analysis of the relation between kinematics and cell activity. Figure 4 depicts the activity of one bimanual-related MI neuron. Trials in which unimanual and bimanual movements were similar are shown in the top displays; trials in which the movements were not similar are shown underneath. Figure 4 shows that selection of trials with similar trajectories in unimanual and bimanual conditions did not lessen the bimanual-related effect, and selection of trials with different trajectories did not increase it. Moreover, the temporal pattern of the neuron's activity is unaffected by the selection of trials. An

analysis of variance (ANOVA) performed on the activity of the cell thus separated showed a highly significant difference between 'unimanual left' and 'bimanual parallel' ($P < 0.001$), and no significant effect of either the selection of trials by movement path (comparing Fig. 4a and b) or the interaction between the variables.

We performed similar analyses on velocity profiles and EMG. For the EMG, we analysed only the four muscles recorded simultaneously with neurons. For each of the three movement variables, we performed this analysis on 16 bimanually related neurons. In no case did the ANOVA indicate a significant effect on the bimanual-related activity, indicating that the bimanual components of many cells do not depend on precise kinematics or muscular activity.

Our results are in agreement with earlier studies in finding contralateral preference in MI, and also in finding significant, albeit weaker and less frequent, ipsilateral activation of MI¹⁰⁻¹⁵. However, the discovery of bimanual-related activity in MI undermines the hypothesis that task-related activity in MI exclusively reflects movement of either limb alone. Indeed, the data suggest that information relevant to coordination effects neural activity of both MI and SMA.

We support the conclusions of previous work^{13,14} examining the bimanual-related activity of single neurons in MI and SMA, which found that a far greater proportion of units in the SMA respond to ipsilateral movements, and that a significant proportion of SMA neurons fire specifically in the bimanual task. We extend these findings to different types of bimanual movements, and demonstrate that neurons can be activated in relation to the specific nature of the bimanual coordination (Fig. 2).

One striking difference exists between our data and the previous work. We find that bimanual-related responses occur at least as frequently in MI as in SMA, whereas the previous work found that a very small fraction (2%) of MI neurons responded exclusively to the unimanual or exclusively to the bimanual movements. This difference may lie in the fact that our monkeys used their entire arm.

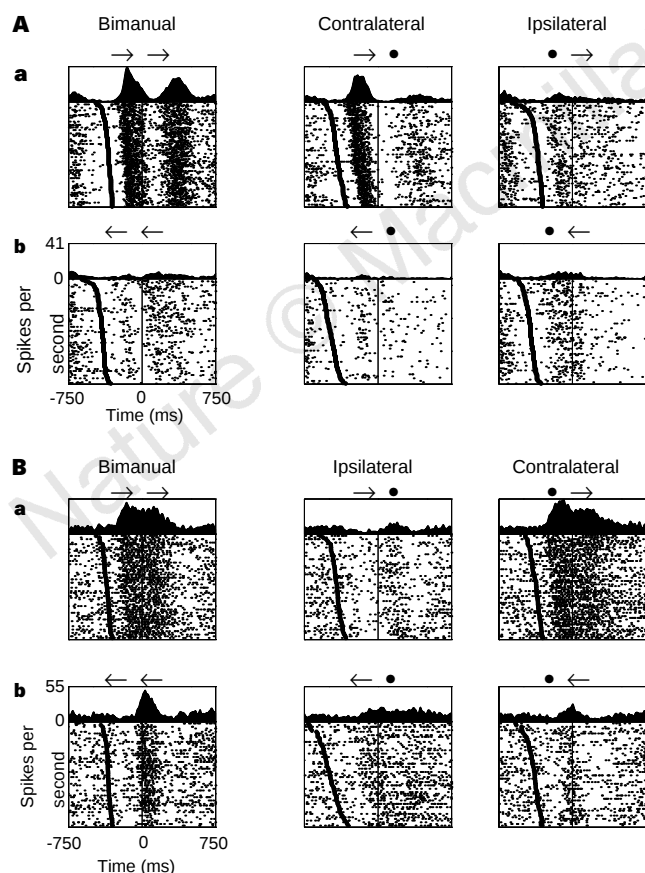


Figure 3 Two MI cells with bimanual-related activity. As Fig. 2, but with only two rows depicting two of the four types of bimanual trials. The movement directions were 45° ($\rightarrow$) and 225° ($\nwarrow$) for both cells. **A**, A cell with directionally selective contralateral activity (**a**, **b**, middle column) and an additional bimanual-related component (**a**, **b**, left column); $n = 159$. **B**, A cell with directionally selective contralateral activity (**a**, **b**, right column) exhibiting a weak bimanual-related component, at the limit of significance. The bimanual-related component appears only when the contralateral arm moves in the non-preferred direction (**b**, left column); $n = 62$.

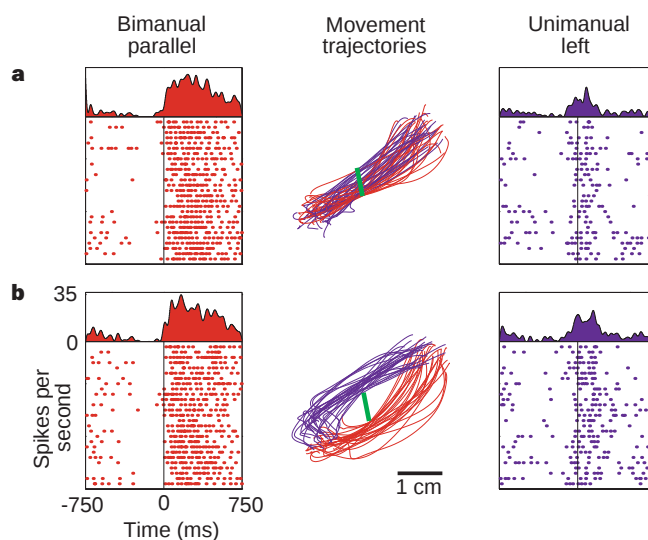


Figure 4 Effect of kinematics on cell response. The raster displays and PSTHs illustrate the activity from a right MI cell in four conditions. The activity of the cell during bimanual parallel movements is on the left (red), and during unimanual left movements is on the right (blue). Middle plots show the movement paths of the left hand for bimanual parallel (red) and unimanual left (blue) movements. **a**, Trials in which the movement path passed through a narrow band (thick green line) located between the origin and target. **b**, Trials that did not pass through the band. The green band was placed to maximize the difference between the trajectories in the lower display. PSTHs are centred on the beginning of the movement, and the scale for all PSTHs is the same; $n = 27$. The trajectories begin in the top right and end in the bottom left of the frame.

Table 2 Many cells in MI and SMA have highly significant bimanual-related components

Significance level	MI for monkey F	MI for monkey G	SMA for monkey F
$<10^{-14}$	13 (18%)	12 (27%)	15 (21%)
$<10^{-10}$	22 (31%)	13 (30%)	20 (27%)
$<10^{-6}$	33 (46%)	20 (45%)	32 (44%)
$<10^{-3}$	53 (74%)	27 (61%)	47 (64%)
$<10^{-2}$	60 (83%)	31 (70%)	53 (73%)
Total	72 (100%)	44 (100%)	73 (100%)

The cumulative number (cumulative percentage) of bimanual-related cells for which activity during bimanual movements differed significantly from activity during unimanual movements for different possible criteria of significance (for monkeys F and G). The level of significance used in the text is $P < 10^{-3}$.

Anatomic and lesion work has raised the suggestion¹⁶ that “each half of the brain has full control over arm, hand, and finger movements contralaterally but only controls arm movements ipsilaterally.” Similarly, studies on the callosal projections of MI have shown that the two hand representations are poorly connected, whereas the two arm areas are strongly connected^{4,17–20}. It is possible that MI is wired contralaterally for distal movements and bilaterally for proximal ones, and that our findings are the electrophysiological correlate of this fact.

The differences we found may also be due to differences in muscle mechanics, if the cells that we recorded are primarily related to muscle activation^{21,22}. However, it has been argued that cells in MI are more significantly related to the intended direction of movement than to precise muscle activation patterns^{23–25}. Indeed, our results indicate that the bimanual-related components did not depend on the precise kinematics or muscle activation.

The most likely source of cortical representations of bimanual coordination are the callosal connections between the hemispheres. The importance of callosal connections for bimanual skills has been repeatedly demonstrated^{5,16}, and perhaps temporal correlation between the hemispheres participates in generating the bimanual-related components^{26–28}. Whatever the source of the bimanual-related signal, our results imply that neuronal mechanisms modify the activity of both SMA and MI during bimanual movements. This mechanism might be the basis of interlimb coordination. □

Methods

Animal and behavioural task. We used two female rhesus monkeys (*Macaca mulatta*) (4.5 and 4.2 kg). The animals' care and surgical procedures used were in accordance with *The NIH Guide for the Care and Use of Laboratory Animals* (1996). A screen was located 30 cm from the monkey. Each trial began when the monkey centred both cursors on ‘origins’ (central circles: Fig. 1) and held them in place for 500 ms. For each arm, one of eight target circles (0.8 cm in diameter) appeared 3 cm from the origin. If only one target appeared (a unimanual trial) the monkey moved only the appropriate arm to achieve the target (Fig. 1Ba, b). If two targets appeared (a bimanual trial) the monkey moved both arms, achieving both targets simultaneously (Fig. 1Bc, d). Two classes of bimanual movements were tested in the recording sessions: parallel, in which targets were located in the same direction from the origins (Fig. 1Bc), and opposite, in which targets were separated by 180° (Fig. 1Bd). Bimanual trials required the monkey to begin movement of the arms within 300 ms of each other and to achieve the two targets within 300 ms of each other. On target acquisition, the monkey held both arms still for at least 500 ms. A velocity threshold determined movement initiation (15 mm s⁻¹).

Recording sessions. Recording chambers (27 × 27 mm) were implanted above both the right and left hemispheres under general anaesthesia in aseptic conditions. During recordings, the monkey sat in a primate chair in a dark chamber, and four electrodes were introduced into each hemisphere. The electrode signals were amplified, filtered and sorted (MCP-8000, MSD, Alpha-Omega), and all spike shapes were sampled at 24 kHz for off-line confirmation of spike sorting. We began data collection by testing the units ‘preferred’ directions for unimanual movements of each limb²⁹. We collected data during

performance of bimanual and unimanual trials in two directions separated by 180°, chosen on the basis of the preferred directions of selected units. At the end of each session, we examined the activity of neurons evoked by passive manipulation of the limbs. Finally, we applied intracortical microstimulation (ICMS; 500 ms of 200-μs cathodal pulses at 300 Hz with an intensity of 10–80 μA) to evoke movements.

Recording sites. We followed standard histological procedures to construct surface maps of all penetrations and to support the physiological definition of MI and SMA. The identification of the proximal arm areas of SMA and MI was based on neuronal responses during the task, passive limb movements, the effect of microstimulation, and the histological analysis. No attempt was made to record from the pre-SMA³⁰. In MI we recorded almost exclusively on the bank, not in the depths of the central sulcus. In SMA we recorded almost exclusively on the medial wall.

Data analysis. We selected single neurons for analysis on the basis of our ability to isolate their spikes. Standard raster displays and peristimulus time histograms were examined. We considered cells to be directionally selective during unimanual movements if their activity before and during movement in opposing directions differed significantly. For example, in Fig. 3A, the contralateral activity in a was compared with that in b, demonstrating directional selectivity during contralateral movements. The Mann–Whitney rank statistic, calculated on the trial-by-trial firing rates during an interval of 500 ms starting 250 ms before movement initiation, determined the statistical significance of differences in activity. Repeating the analysis on a 250-ms epoch and a 1-s epoch did not affect our results. The level of significance used here is $P < 0.001$ for all statistical tests. We compared activity during bimanual movements with the corresponding unimanual movement to which the cell responded most strongly. Thus, four Mann–Whitney tests (one for each type of bimanual movement) were performed for each cell. For example, in Fig. 3Aa we compared the bimanual and contralateral responses, and in Fig. 3Ab we compared the bimanual and ipsilateral responses. The significance over the four tests, and the criterion was divided by four to correct for the compounded tests. Our definition of bimanual-related activity is similar to the exclusive cells in previous studies of bimanual activity^{13,14}. Our use of the term bimanual-related does not include cells with pronounced activity during both bimanual and unimanual movements (bilateral cells in previous studies).

EMG recording. We recorded surface EMGs for nine muscles bilaterally, including deltoid, rhomboids, pectoralis major, latissimus dorsi, teres major, biceps brachii, triceps brachii, flexor carpi ulnaris and extensor carpi ulnaris. We recorded two of these muscles (deltoid and flexor carpi ulnaris) bilaterally during neural recording sessions in monkey G.

Received 17 April; accepted 23 June 1998.

- Kelso, J. A. S., Southard, D. L. & Goodman, D. On the coordination of two-handed movements. *J. Exp. Psychol. Hum. Percept. Perform.* **5**, 229–238 (1979).
- Franz, E. A., Zelaznik, H. N. & McCabe, G. Spatial topological constraints in a bimanual task. *Acta Psychol.* **77**, 137–151 (1991).
- Wiesendanger, M. In *Role of the Cerebellum and Basal Ganglia in Voluntary Movement* (eds Mano, N., Hamada, I. & DeLong, M. R.) 253–266 (Elsevier, Amsterdam, 1993).
- Wiesendanger, M., Rouiller, E. M., Kazennikov, O. & Perrig, S. Is the supplementary motor area a bilaterally organized system? *Adv. Neurol.* **70**, 85–93 (1996).
- Brinkman, C. Supplementary motor area of the monkey's cerebral cortex: short- and long-term deficits after unilateral ablation and the effects of subsequent callosal section. *J. Neurosci.* **4**, 918–929 (1984).
- Brinkman, C. & Porter, R. Supplementary motor area in the monkey: activity of neurons during performance of a learned motor task. *J. Neurophysiol.* **42**, 681–709 (1979).
- Karol, E. A. & Pandya, D. N. The distribution of the corpus callosum in the rhesus monkey. *Brain* **94**, 471–486 (1971).
- Tokuno, H. & Tanji, J. Input organization of distal and proximal forelimb areas in the monkey primary motor cortex: a retrograde double labeling study. *J. Comp. Neurol.* **33**, 199–209 (1993).
- Luppino, G., Matelli, M., Camarda, R. & Rizzolatti, G. Corticocortical connections of area F3 (SMA-proper) and area F6 (pre-SMA) in the macaque monkey. *J. Comp. Neurol.* **338**, 114–140 (1993).
- Lecas, J. C., Requin, J., Anger, C. & Vitton, N. Changes in neuronal activity of the monkey precentral cortex during preparation for movement. *J. Neurophysiol.* **56**, 1680–1702 (1986).
- Crammond, D. J. & Kalaska, J. F. Preparatory neuronal activity in premotor cortex during an instructed-delay period: relation to contra and ipsilateral arm movements. *Soc. Neurosci. Abstr.* **17**, 308–308 (1991).
- Perepelkin, P. D. & Schwartz, A. B. Simultaneous populations of single-cell activity recorded bilaterally in primate motor cortex. *Soc. Neurosci. Abstr.* **22**, 2022–2022 (1996).
- Tanji, J., Okano, K. & Sato, K. C. Relation of neurons in the nonprimary motor cortex to bilateral hand movement. *Nature* **327**, 618–620 (1987).
- Tanji, J., Kazuhiko, O. & Kazuko, C. S. Neuronal activity in cortical motor areas related to ipsilateral, contralateral, and bilateral digit movements of the monkey. *J. Neurophysiol.* **60**, 325–343 (1988).
- Kim, S. G. et al. Functional magnetic resonance imaging of motor cortex: hemispheric asymmetry and handedness. *Science* **261**, 615–617 (1993).
- Brinkman, J. & Kuypers, G. J. M. Cerebral control of contralateral and ipsilateral arm, hand and finger movements in the split-brain rhesus monkey. *Brain* **96**, 653–674 (1973).

17. Jenny, A. B. Commissural projections of the cortical hand motor area in monkeys. *J. Comp. Neurol.* **188**, 137–145 (1979).
18. Muakkassa, K. F. & Strick, L. Frontal lobe inputs to primate motor cortex: evidence for four somatotopically organized 'premotor' areas. *Brain Res.* **177**, 176–182 (1979).
19. Jones, E. G. In *Cerebral Cortex* (eds Jones, E. G. & Peters, A.) 113–183 (Plenum, New York, 1986).
20. Rouiller, E. M. *et al.* Transcallosal connections of the distal forelimb representations of the primary and supplementary motor cortical areas in macaque monkeys. *Exp. Brain Res.* **102**, 227–243 (1994).
21. Evars, E. V. Relation of pyramidal tract activity to force exerted during voluntary movement. *J. Neurophysiol.* **31**, 14–27 (1968).
22. Scott, S. H. Comparison of onset time and magnitude of activity for proximal arm muscles and motor cortical cells before reaching movements. *J. Neurophysiol.* **77**, 1016–1022 (1997).
23. Georgopoulos, A. P., Ashe, J., Smyrnis, N. & Taira, M. The motor cortex and the coding of force. *Science* **256**, 1692–1695 (1992).
24. Schwartz, A. B. Direct cortical representation of drawing. *Science* **265**, 540–542 (1994).
25. Taira, M., Boline, J., Smyrnis, N., Georgopoulos, A. P. & Ashe, J. On the relations between single cell activity in the motor cortex and the direction and magnitude of three-dimensional static isometric force. *Exp. Brain Res.* **109**, 367–376 (1996).
26. Donchin, O., Gribova, A., Bergman, H. & Vaadia, E. How do the two hemispheres of the cortex communicate in coordinating the actions of the arms: single unit study in a behaving monkey. *Soc. Neurosci. Abstr.* **22**, 2022–2022 (1996).
27. Roelfsema, P. R., Engel, A. K., König, P. & Singer, W. Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature* **385**, 157–161 (1997).
28. Murthy, V. N. & Fetz, E. E. Synchronization of neurons during local field potential oscillations in sensorimotor cortex of awake monkeys. *J. Neurophysiol.* **76**, 3968–3982 (1996).
29. Georgopoulos, A. P., Kalaska, J. F., Caminiti, R. & Massey, J. T. On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex. *J. Neurosci.* **2**, 1527–1537 (1982).
30. Matsuzaka, Y., Aizawa, H. & Tanji, J. A motor area rostral to the supplementary motor area (presupplementary motor area) in the monkey: neuronal activity during a learned motor task. *J. Neurophysiol.* **68**, 653–662 (1992).

Acknowledgements. We thank S. P. Wise and J. Kalaska for their comments on earlier versions of the manuscript; A. P. Mitz for contributions to the style and content of the work; Y. Donchin for help during the surgical procedures; and the Clure Foundation for the fellowship that supported O.D. throughout this project. This research is supported in part by a grant from the US-Israel Binational Foundation and by a grant from the Israel Science Foundation.

Correspondence and requests for materials should be addressed to E.V. (e-mail: eilon@hbf.huji.ac.il).

Early-blind human subjects localize sound sources better than sighted subjects

N. Lessard*, M. Paré†, F. Lepore*† & M. Lassonde*†

* Groupe de Recherche en Neuropsychologie Expérimentale, Département de Psychologie, and † Centre de Recherche en Sciences Neurologiques, Université de Montréal, C.P. 6128, Succ. Centre-ville, Montréal, Québec H3C 3J7, Canada

Do blind persons develop capacities of their remaining senses that exceed those of sighted individuals? Besides anecdotal suggestions, two views based on experimental studies have been advanced¹. The first proposes that blind individuals should be severely impaired, given that vision is essential to develop spatial concepts². The second suggests that compensation occurs through the remaining senses, allowing them to develop an accurate concept of space³. Here we investigate how an ecologically critical function, namely three-dimensional spatial mapping, is carried out by early-blind individuals with or without residual vision. Subjects were tested under monaural and binaural listening conditions. We find that early-blind subjects can map the auditory environment with equal or better accuracy than sighted subjects. Furthermore, unlike sighted subjects, they can correctly localize sounds monaurally. Surprisingly, blind individuals with residual peripheral vision localized sounds less precisely than sighted or totally blind subjects, confirming that compensation varies according to the aetiology and extent of blindness⁴. Our results resolve a long-standing controversy in that they provide behavioural evidence that totally blind individuals have better auditory ability than sighted subjects, enabling them to compensate for their loss of vision.

We examined how subjects with congenital deficits affecting the peripheral visual system (retina and optic nerve) localized sounds in space. Four groups were tested: totally blind subjects ($n = 8$); blind subjects with residual vision in the peripheral field ($n = 3$); nor-

mally sighted but blindfolded controls ($n = 7$); and sighted controls ($n = 29$). Subjects were asked to localize a sound source presented on the horizontal plane. The sounds were delivered randomly through 16 loudspeakers mounted on a semicircular perimeter. Peri-central field was defined arbitrarily as the space covered by the four centrally located loudspeakers (up to 16° on either side of the midline), whereas the lateral fields extended to 78°. Subjects were tested under monaural and binaural conditions, each providing a specific set of cues to sound localization. Binaural cues refer to the discrepancies of inputs between the ears in terms of timing and intensity, whereas monaural cues arise from the spectral filtering of sounds by the circonvolution of the pinna.

The sound-localization performance in the binaural condition is plotted in Fig. 1. Because a preliminary analysis showed that the mean error scores in each field were not different for blindfolded and sighted controls ($P > 0.05$), their results were pooled (Fig. 1a). The principal results indicate that totally blind subjects were at least

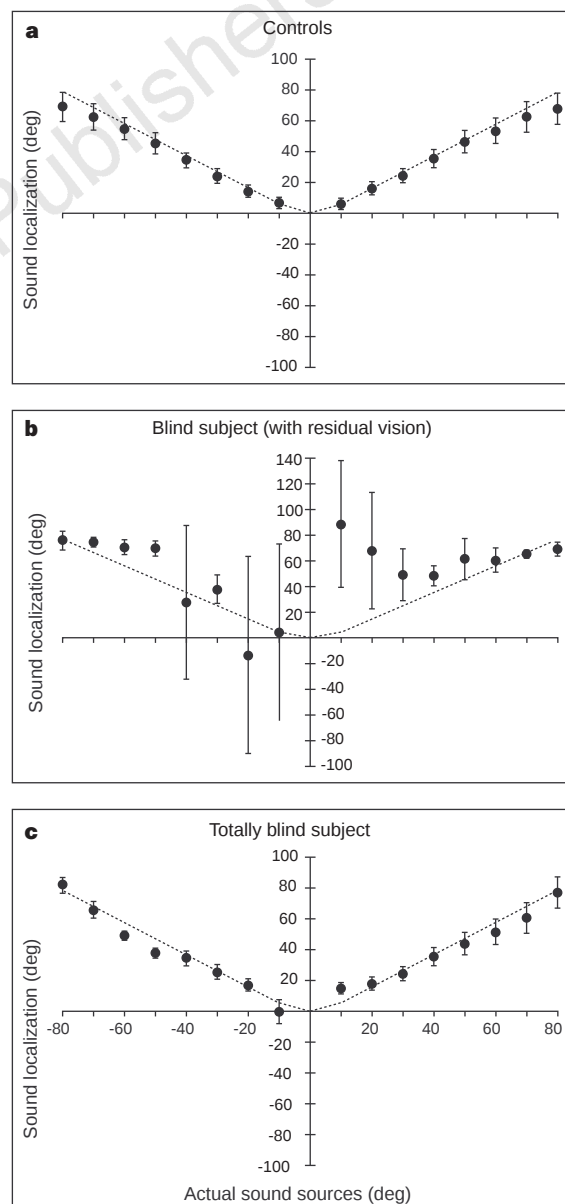


Figure 1 Binaural sound localization performance obtained in each hemifield. **a**, Controls ($n = 36$); **b**, a blind subject with residual vision; and **c**, one representative early-blind subject. The dashed lines indicate the actual sound source, whereas the dots refer to the perceived target locations with their respective standard deviations.

as accurate as sighted controls (Fig. 1c), although because of a ceiling effect it was not possible to show that they were actually better than the latter. To bring out these differences, it may be necessary to render the task more difficult by using, for example, narrow-band stimuli. The results nonetheless demonstrate that blind individuals can develop a three-dimensional map of space using auditory information.

This raises questions regarding the role of vision in calibrating auditory space. Experience-dependent plasticity in the visual calibration of the auditory space map has been demonstrated in the barn owl's optic tectum⁵, where neurons form a map of auditory space based on interaural time and level differences. When visual and auditory worlds are experimentally misaligned by rearing owls with prismatic spectacles that displace the visual field in azimuth, tectal neurons acquire responses to interaural time differences that correspond to the displaced visual fields and the neurons abandon responses to normal interaural time differences. This seems to support the notion that vision is essential for calibrating space, including auditory space. Our results, however, indicate clearly that this is not the case with humans who totally lack the sense of vision, as our early-blind subjects develop excellent spatial representation in the absence of this modality.

Blind subjects with residual vision were less accurate than all other subjects, particularly in the pericentral field (Fig. 1b). This result was unexpected as it had been predicted that these subjects would show normal localization behaviour in peripheral fields (where vision was present), and a performance similar to that of the early-blind subjects in central visual field (where vision was lacking). Their subnormal performance across the entire auditory field has three possible explanations. First, these subjects would need to develop an auditory map of space in part supported by vision (peripheral field) and in part independent of vision (central field), which might cause some confusion. Second, if auditory compensation in blind subjects depends on the recruitment of the deafferented sensory areas (see below), the latter would not show a similar amount of plasticity if they were still stimulated, albeit at a reduced rate, by their normal afferences. Third, these partially blind subjects demonstrated abnormal orienting behaviours, which might interfere with correct localization, as they often would fixate the source of a sound (the voice of the experimenter, the test sound during pretest, when head movement was allowed) by turning their head so that the source would be visible by their residual visual field.

In the monaural condition, where the stimuli were presented while one ear was blocked, sighted and blindfolded controls were once again indistinguishable from each other ($P > 0.05$) and their results were pooled. Figure 2a shows that these subjects localized the sound on the side of the unobstructed ear, approximately at mid-field, independently of the side from which it originated. As in the binaural condition, the performance of blind subjects with residual vision was worse than that of the others (Fig. 2b). They also showed the positional bias in favour of the unobstructed ear when localizing a sound presented on the side of the obstructed ear. The performance of the totally blind subjects, however, was quite exceptional. Half (Fig. 2c) appeared to respond like controls, that is, with a positional bias favouring the side of the unobstructed ear. However, contrary to controls, their performance showed more variability when the stimulus was presented on the side of the obstructed ear. Moreover, they often reported that the sound presented on this side seemed qualitatively different. For example, one subject reported that the sound seemed to show a narrower bandpass, as if some frequencies were cut off. The most interesting aspect of this set of results, as shown in Fig. 2d, was the fact that the other four totally blind subjects localized the sound on the appropriate side—that is, even when it was presented on the side of the obstructed ear.

The ability to localize the sound in the correct hemifield, irrespective of accuracy, in both the monaural and binaural testing

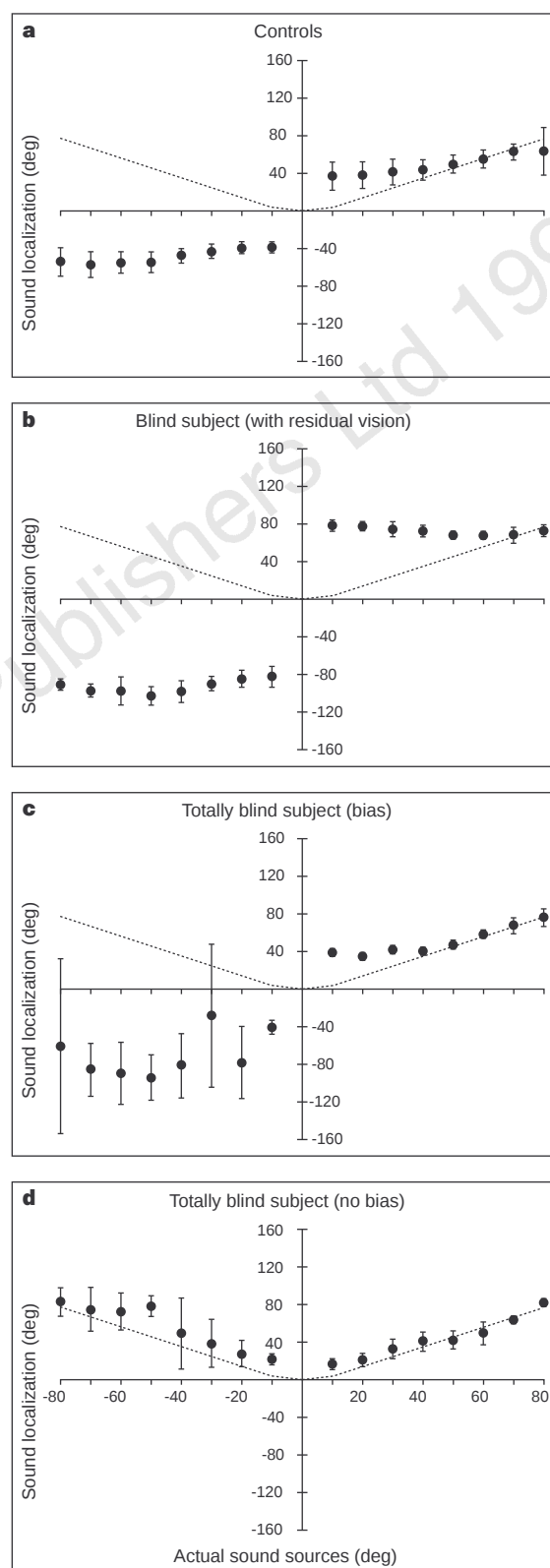


Figure 2 Monaural sound localization performance obtained in one hemifield. **a**, Controls ($n = 36$); **b**, one representative blind subject with residual vision; **c**, one representative totally blind subject with a directional bias, and **d**, one totally blind subject who correctly localized the sound with no directional bias. Results are shown for one monaural condition only (left-ear-obstructed) as the other condition (right-ear-obstructed) yielded identical results. The dashed lines indicate the actual sound sources, whereas the dots refer to the perceived target locations with their respective standard deviations.

conditions is illustrated in Fig. 3. The most striking observation concerns the totally blind subjects: half of them localized correctly the sound presented on the side of the obstructed ear. This behaviour, moreover, was neither marginal, as performance was nearly perfect, nor was it observed in any of the large number of sighted subjects (29 normally stimulated, 7 tested while blind-folded). This finding, coupled with the observation that the other four totally blind subjects, although unable to localize on the correct side, did report qualitative differences in the sound presented to the blocked ear—again a behaviour not seen in any of the other groups—suggests that blind individuals utilize monaural cues more efficiently to auditorily explore their environment.

The finding that early-blind subjects can perform better than sighted subjects can be attributed to reorganizations in neuronal populations involved in processing localization cues and/or to improved learning. Thus, compensation may result from the increased use of spectral information within or between the structures normally involved in spatial analysis, including the superior^{6–8} and inferior colliculus⁹, as well as primary auditory cortex^{10,11}. Alternately, compensation may occur through the recruitment of brain structures left unused by the lack of visual input. In cats, visual deprivation beginning shortly after birth results in compensatory effects at the collicular level¹², improved auditory responses in neurons involved in visual processing¹³, and superior localization performance¹⁴. Similarly, in early-blind humans, Kujala and colleagues^{15,16} showed that, using event-related potentials or magnetoencephalography, the response to auditory stimuli was distributed more posteriorly compared with sighted controls, suggesting increased use of parietal or perhaps even occipital brain areas in sound localization. They propose that this cross-modal reorganization might rest on the unmasking¹⁷ of latent, pre-existing auditory connections to the visual cortex. In addition, better learning strategies could explain their superior performance, as psychophysical studies have shown that normal subjects fitted with an earplug, who demonstrate a prominent displacement in their localization judgement towards the side of the open ear, can increase their precision with learning¹⁸. Similarly, Slattery and Middlebrooks¹⁹ studied sound localization in five unilaterally deaf subjects and showed that three of them had little or no localization bias towards the functional ear. This implies the development of

strategies for localizing sounds monaurally, which might likewise be adopted by blind persons. □

Methods

The apparatus used to test sound localization, described in detail elsewhere²⁰, consisted of 16 loudspeakers mounted on a graduated semicircular perimeter (radius 50 cm) positioned at $\pm 5^\circ$, $\pm 16^\circ$, $\pm 26^\circ$, $\pm 37^\circ$, $\pm 47^\circ$, $\pm 58^\circ$, $\pm 68^\circ$ and $\pm 78^\circ$ on either side of the mid-sagittal plane. The subject was seated in the centre of the perimeter, the head placed on a head-rest attached to the chair, and the speakers were positioned at ear level. All testing was done in an anechoic chamber. The stimuli were two broad-band noise bursts that lasted 30 ms (10-ms rise and 10-ms fall time) with an interburst interval of 30 ms (total stimulus duration 90 ms). The sound pressure level (SPL) was maintained at 40 dB SPL (20 μ Pa). The stimulus was delivered through a randomly selected loudspeaker and repeated five times for each position. The head was not restrained other than being apposed on the head-rest, but a warning buzzer told the subjects that a stimulus was about to be presented and that they should maintain stable head position and fixate straight ahead. Only a subgroup of control subjects were blindfolded; the others, including the partly blind subjects, were free to have their eyes open. Compliance to all instructions was verified by an experimenter who stayed in the chamber and remained behind the subject. The speaker array was covered with black cloth so that sighted subjects could not see the actual speakers and response consisted of pointing with the dominant hand to the apparent source of stimulation. Lines graduated in 1° steps were drawn on the perimeter and the experimenter recorded the response of the subject. For monaural testing, either the right or the left ear was plugged with a soft foam earplug (mean attenuation = 37.5 dB SPL) and covered by a hearing protector muff (mean attenuation, 29 dB SPL). To ensure that no sounds could be perceived with the earplug and hearing protector muff in place, a preliminary study was done to test the efficacy of the earplugs during which both ears were blocked. Broad-band noise bursts ranging from 25 to 60 dB SPL were delivered randomly from four different positions ($\pm 16^\circ$ and $\pm 68^\circ$) for a total of 20 presentations. No subject reported hearing a stimulus presented within these ranges.

Received 23 February; accepted 19 June 1998.

1. Rauschecker, J. P. Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends Neurosci.* **18**, 36–43 (1995).
2. Axelrod, S. *Effects of Early Blindness* (Am. Found. Blind, New York, 1959).
3. Rice, C. E. Early blindness, early experience and perceptual enhancement. *Res. Bull. Am. Found. Blind* **22**, 1–22 (1970).
4. Kujala, T. et al. Electrophysiological evidence for cross-modal plasticity in humans with early- and late-onset blindness. *Psychophysiology* **34**, 213–216 (1997).
5. Feldman, D. E. & Knudsen, E. I. An anatomical basis of visual calibration of the auditory space map in the barn owl's midbrain. *J. Neurosci.* **17**, 6820–6837 (1997).
6. Chalupa, L. M. & Rhoades, R. W. Responses of visual, somatosensory, and auditory neurones in the golden hamster's superior colliculus. *J. Physiol.* **270**, 595–626 (1977).
7. King, A. J. A map of auditory space in the mammalian brain: neural computation and development. *Exp. Physiol.* **78**, 559–590 (1993).
8. Middlebrooks, J. C. & Knudsen, E. I. A neural code for auditory space in the cat's superior colliculus. *J. Neurosci.* **4**, 2621–2634 (1984).
9. Spitzer, M. W. & Semple, M. N. Interaural phase coding in auditory midbrain: influence of dynamic stimulus features. *Science* **254**, 721–723 (1991).
10. Stumpf, E., Toronchuk, J. M. & Cynader, M. S. Neurons in cat primary auditory cortex sensitive to correlates of auditory motion in three dimensional space. *Exp. Brain Res.* **88**, 158–168 (1992).
11. Toronchuk, J. M., Stumpf, E. & Cynader, M. S. Auditory cortex neurons sensitive to correlates of auditory motion—underlying mechanisms. *Exp. Brain Res.* **88**, 169–180 (1992).
12. Rauschecker, J. P. & Harris, L. R. Auditory compensation of the effects of visual deprivation in the cat's superior colliculus. *Exp. Brain Res.* **50**, 69–83 (1983).
13. Rauschecker, J. P. & Korte, M. J. Auditory compensation for early blindness in cat cerebral cortex. *J. Neurosci.* **13**, 4538–4548 (1993).
14. Rauschecker, J. P. & Knipert, U. Auditory localization behaviour in visually deprived cats. *Eur. J. Neurosci.* **6**, 149–160 (1994).
15. Kujala, T., Alho, K., Paavilainen, P., Summala, H. & Näätänen, R. Neural plasticity in processing of sound localization by the early blind: an event-related potential study. *Electroencephalogr. Clin. Neurophysiol.* **84**, 469–472 (1992).
16. Kujala, T. et al. Auditory an somatosensory event-related potentials in early blind humans. *Exp. Brain Res.* **104**, 519–526 (1995).
17. Jacobs, K. M. & Donoghue, J. P. Reshaping the cortical motor map by unmasking latent intracortical connections. *Science* **251**, 944–947 (1991).
18. Butler, R. A. An analysis of the monaural displacement of sound in space. *Percept. Psychophys.* **41**, 1–7 (1987).
19. Slattery, W. H. & Middlebrooks, J. C. Monaural sound localization: acute versus chronic unilateral impairment. *Hearing Res.* **75**, 38–46 (1994).
20. Poirier, P., Miljourns, S., Lassonde, M. & Lepore, F. Sound localization in acallosal human listeners. *Brain* **116**, 53–69 (1993).

Acknowledgements. We thank the Regroupement pour les Aveugles et Amblyopes de Montréal (RAAM), especially S. Poulin and F. Boulet, for their assistance in recruiting participants. This work was supported by grants from NSERC and FCAR.

Correspondence should be addressed to F.L. (e-mail: lepore@ere.umontreal.ca).

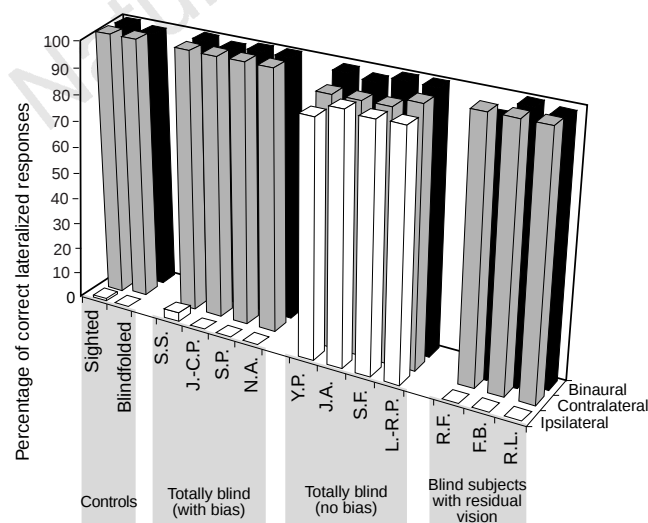


Figure 3 Percentage of trials in which the sounds were correctly lateralized in the binaural condition and in each of the monaural conditions when the sound was presented ipsilaterally and contralaterally to the obstructed ear: control groups; four totally blind subjects who localized the sound with a positional bias similar to that found in the controls; four totally blind subjects who correctly localized the sound; and three blind subjects with residual vision.

Impaired febrile response in mice lacking the prostaglandin E receptor subtype EP₃

Fumitaka Ushikubi^{*,†}, Eri Segi^{†,‡}, Yukihiro Sugimoto[†], Takahiko Murata^{*}, Toshiyuki Matsuoka^{*}, Takuya Kobayashi^{*}, Hiroko Hizaki[†], Kazuhito Tuboi[†], Masato Katsuyama[†], Atsushi Ichikawa[†], Takashi Tanaka[§], Nobuaki Yoshida[§] & Shuh Narumiya^{*}

^{*} Department of Pharmacology, Faculty of Medicine, and [†] Department of Physiological Chemistry, Faculty of Pharmaceutical Sciences, Kyoto University, Kyoto 606-8315, Japan

[§] Division of Molecular and Cellular Immunology, Research Institute, Osaka Medical Center for Maternal and Child Health, Osaka 594-1101, Japan

[‡] These authors contributed equally to this work.

Fever, a hallmark of disease, is elicited by exogenous pyrogens, that is, cellular components, such as lipopolysaccharide (LPS), of infectious organisms, as well as by non-infectious inflammatory insults. Both stimulate the production of cytokines, such as interleukin (IL)-1 β , that act on the brain as endogenous pyrogens¹. Fever can be suppressed by aspirin-like anti-inflammatory drugs. As these drugs share the ability to inhibit prostaglandin biosynthesis², it is thought that a prostaglandin is important in fever generation. Prostaglandin E₂ (PGE₂) may be a neural mediator of fever³, but this has been much debated^{4,7}. PGE₂ acts by interacting with four subtypes of PGE receptor, the EP₁, EP₂, EP₃ and EP₄ receptors⁸. Here we generate mice lacking each of these receptors by homologous recombination. Only mice

lacking the EP₃ receptor fail to show a febrile response to PGE₂ and to either IL-1 β or LPS. Our results establish that PGE₂ mediates fever generation in response to both exogenous and endogenous pyrogens by acting at the EP₃ receptor.

The mouse genes encoding the EP₁ and EP₃ receptors were individually disrupted (Fig. 1a, b) and chimaeric mice were generated. These animals were backcrossed with C57BL/6 mice, and the resulting heterozygous littermates (EP₁^{+/-} or EP₃^{+/-}) were bred to produce homozygous EP₁^{-/-} or EP₃^{-/-} mice (Fig. 1c, d). Homozygous mice were born at the predicted Mendelian frequency, grew normally, lived longer than 1 year and were fertile. We detected no apparent morphological abnormalities. We used the wild-type and homozygous mutant mice of these F₂ progeny for subsequent analysis, except in experiments using LPS. Targeting of the EP₄ receptor gene has been described⁹ and the generation of EP₂-receptor-deficient mice is described in Supplementary information. Disruption of the EP₁ and EP₃ receptor genes was verified by northern blot hybridization, which failed to detect messenger RNAs encoding the respective receptors in EP₁^{-/-} and EP₃^{-/-} mice (Fig. 1e, f). In addition, EP₃-receptor-specific binding of [³H]PGE₂ was not detected in crude kidney membranes of EP₃^{-/-} mice, whereas a value of 65 $\pm$ 2 fmol per mg protein was obtained for such binding in wild-type mice (Fig. 1g).

Although PGE₂ produces fever when injected intracerebroventricularly (i.c.v.)⁶, the receptor subtype that mediates this action has not been determined. We therefore addressed this issue using mice lacking each of the four receptor subtypes. We first confirmed that PGE₂ induces fever in wild-type mice of the F₂ progeny. Wild-type animals developed a fever with a maximal increase in body temperature of about 2 °C apparent at 20–25 min after i.c.v. administration of 1 nmol PGE₂ (Fig. 2a). Wild-type mice injected with vehicle showed a transient reduction in body temperature due to the ether anaesthesia. PGE₂ also induced fever in EP₁^{-/-}, EP₂^{-/-} and

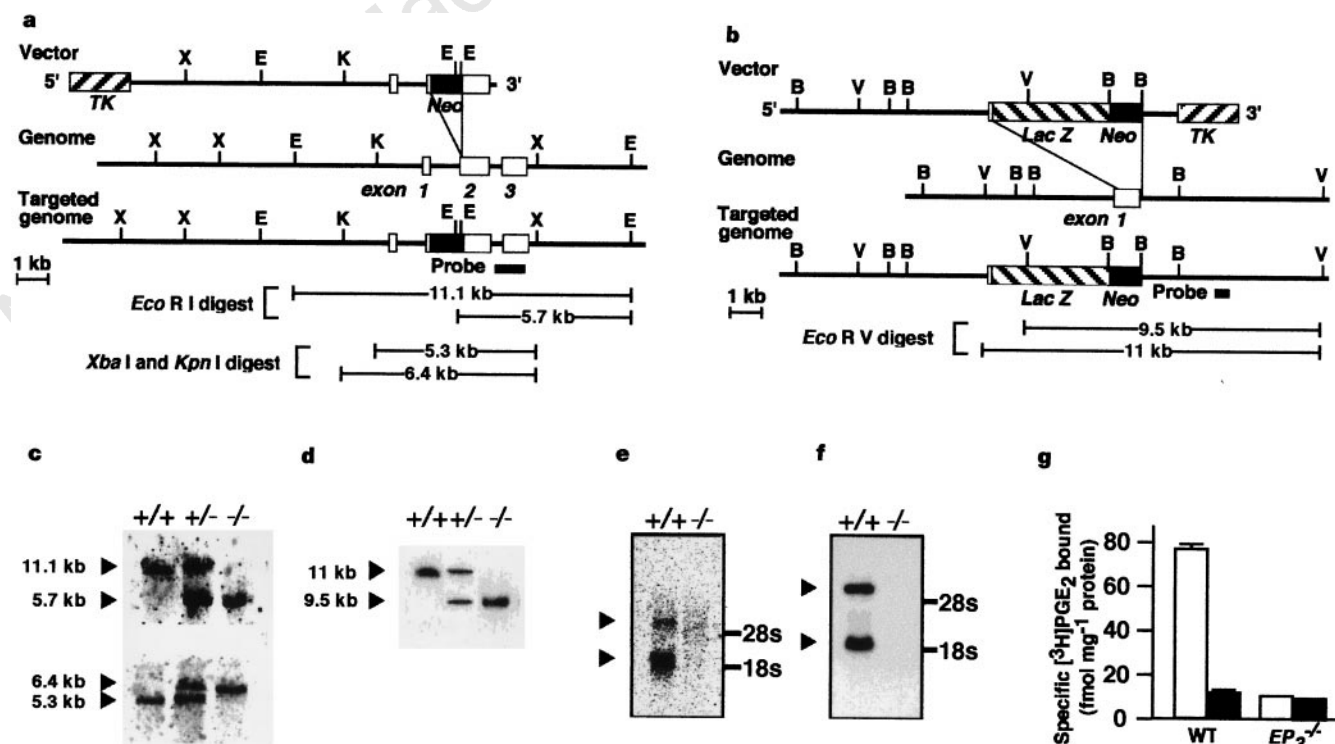


Figure 1 Targeted disruption of the EP₁ and EP₃ receptor genes. **a**, **b**, The targeting vectors, organization of the genes, and structures of the targeted genomes are shown for the EP₁ (**a**) and EP₃ (**b**) receptors. Restriction sites indicated: X, XbaI; E, EcoRI; K, KpnI; B, BamHI; V, EcoRV. Probes used in Southern hybridization are indicated, with the sizes of hybridizing fragments shown below. **c**, **d**, Southern blot analysis of genomic DNA from EP₁ (**c**) and EP₃ (**d**) receptor

mutants. Genomic DNA digested with **c**, either EcoRI (above) or XbaI plus KpnI (below), **d**, with EcoRV was hybridized with the respective external probe. **e**, **f**, Northern blot analysis. +/+, wild type; -/-, EP₁^{-/-} (**e**) or EP₃^{-/-} (**f**) mice. Arrowheads indicate the positions of specific transcripts. **g**, Loss of binding to the EP₃ in EP₃^{-/-} mice. Open columns, specific [³H]PGE₂ binding; filled columns, specific [³H]PGE₂ binding in the presence of M&B28767, an EP₃ receptor agonist.

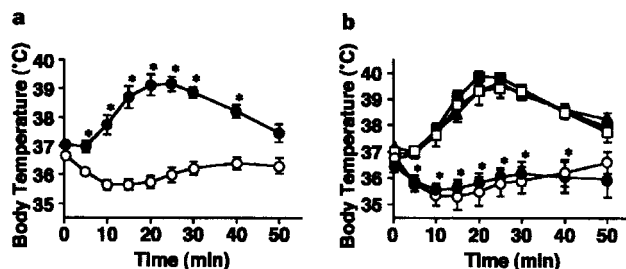


Figure 2 Effects of PGE₂ on body temperature in wild-type and receptor-deficient mice. **a**, PGE₂-induced fever in wild-type mice. Animals were injected i.c.v. with either PGE₂ (filled circles, *n* = 9) or vehicle (open circles, *n* = 9), and rectal temperature was monitored. Asterisk indicates *P* < 0.01 versus vehicle control. **b**, Selective loss of PGE₂-induced febrile response in EP₃^{-/-} mice. PGE₂ was injected i.c.v. into EP₁^{-/-} (filled squares, *n* = 5), EP₂^{-/-} (open squares, *n* = 5), EP₃^{-/-} (filled circles, *n* = 5) and EP₄^{-/-} (filled triangles, *n* = 5) mice. Vehicle was injected i.c.v. into EP₃^{-/-} mice (open circles, *n* = 9). Asterisk indicates *P* < 0.01 for EP₃^{-/-} mice versus wild-type mice injected with PGE₂.

EP₄^{-/-} mice; the extent and time course of these fevers were similar to those observed in wild-type mice (Fig. 2b). The injection of vehicle did not induce fever in these mice (data not shown). In contrast, EP₃^{-/-} mice failed to respond to PGE₂; the body temperature of EP₃^{-/-} animals injected with PGE₂ did not differ significantly from that of those injected with vehicle (Fig. 2b). These results thus show that EP₃ receptors mediate the febrile response to PGE₂.

As most EP₄^{-/-} mice die in the neonatal period, as described^{9,10}, and therefore the number of adult EP₄^{-/-} mice available was limited, we had to exclude these animals from subsequent analysis.

We next examined the febrile response to IL-1β, a representative endogenous pyrogen¹. IL-1β administered intravenously (i.v.) into wild-type mice induced a transient fever that peaked 20 min after the injection and gradually decreased thereafter (Fig. 3a). Both EP₁^{-/-} and EP₂^{-/-} mice showed a similar febrile response to i.v. injection of IL-1β, whereas no fever induction was apparent in EP₃^{-/-} mice (Fig. 3a). These results indicate that IL-1β injected i.v. may stimulate the production of PGE₂, which then acts at EP₃ receptors to induce fever.

IL-1β has often been injected i.c.v. to analyse the biochemical steps that result in fever generation. I.c.v. injection of IL-1β has been suggested to induce fever by a mechanism that differs from that produced by IL-1β injected i.v., and the former mechanism may not involve PGE₂. We therefore studied the role of EP₃ receptors in fever generated in response to i.c.v. injection of IL-1β. As shown for other species^{4,11}, IL-1β administered i.c.v. into wild-type mice induced fever with a late onset, slower rate of increase and longer duration than that observed with i.v. injection, indicating that IL-1β reached its site of action in the brain more rapidly when injected i.v. than when injected i.c.v. (Fig. 3b). EP₁^{-/-} and EP₂^{-/-} mice showed a febrile response to IL-1β injected i.c.v. that was similar to that of the wild-type mice, whereas EP₃^{-/-} mice again did not respond differently to IL-1β and vehicle (Fig. 3c). These results indicate that the EP₃ receptor mediates the IL-1β-induced febrile response, irrespective of the route of administration of IL-1β.

Fever generated by exogenous pyrogens such as LPS cannot be

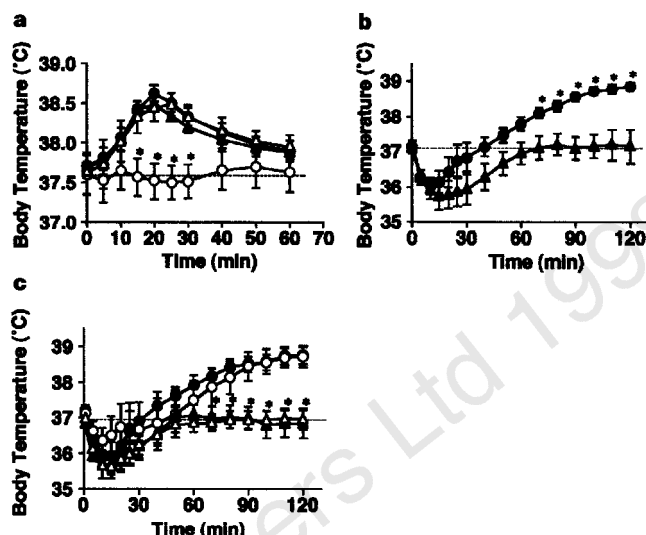


Figure 3 Loss of the IL-1β-induced febrile response in EP₃-receptor-deficient mice. **a**, IL-1β (100 ng) was injected i.v. into wild-type (filled circles, *n* = 9), EP₁^{-/-} (filled triangles, *n* = 6), EP₂^{-/-} (open triangles, *n* = 6) and EP₃^{-/-} mice (open circles, *n* = 6). Asterisk indicates *P* < 0.01 versus wild-type mice. **b**, Either IL-1β (10 ng) (filled circles, *n* = 14) or vehicle (filled triangles, *n* = 9) was injected i.c.v. into wild-type mice. Asterisk indicates *P* < 0.01 versus vehicle control. **c**, IL-1β was injected i.c.v. into EP₁^{-/-} (filled circles, *n* = 5), EP₂^{-/-} (open circles, *n* = 6) and EP₃^{-/-} mice (filled triangles, *n* = 8); other EP₃^{-/-} mice were injected with vehicle (open triangles, *n* = 9). Asterisk indicates *P* < 0.01 EP₃^{-/-} mice versus wild-type mice injected with IL-1β i.c.v. Dotted lines indicate the body temperature before injection.

accounted for by the action of IL-1β alone^{12,13}. LPS, for example, induces the production of other endogenous pyrogens, including IL-6, tumour necrosis factor, macrophage inflammatory protein-1 and IL-8 (ref. 1), the latter two of which may induce fever in a postaglandin-independent manner^{14,15}. It has also been suggested that LPS may act directly on the brain to generate fever under some conditions¹⁶. We therefore determined whether the EP₃ receptor also contributes to LPS-induced fever. Responsiveness to LPS differs among different mouse strains¹⁷. Because the wild-type mice of the F₂ progeny, with a mixed genetic background of 129/Sv and C57BL/6 strains, showed a variable febrile response to LPS, we backcrossed EP₁^{-/-} or EP₃-deficient mice to C57BL/6 mice four times, and used the resulting N4 mice for these experiments. We first stimulated peritoneal macrophages isolated from these N4 mice with LPS and examined their responsiveness to LPS by determining the production of IL-1β and IL-6 (ref. 18). Macrophages from both EP₁^{-/-} and EP₃^{-/-} N4 mice produced the two cytokines in amounts similar to those observed with cells from the wild-type mice (Fig. 4a), indicating that the initial step of the response to LPS may not be impaired in these N4 animals. We next analysed the febrile response to LPS in these mice. LPS (10 mg kg⁻¹) injected i.v. into wild-type mice induced a fever that peaked at +1.0°C at 20–25 min after injection of LPS (Fig. 4b). EP₁^{-/-} mice showed a response to LPS similar to that of the wild-type animals, whereas the body temperature of EP₃^{-/-} mice injected with LPS remained similar to that of wild-type mice injected with vehicle (Fig. 4b). These results show that EP₃ receptors are essential in the febrile response to LPS, and that prostaglandin-independent mechanisms do not contribute significantly, if at all, to fever generation, at least in this model. Finally, we confirmed that EP₃-deficient mice were able to develop a febrile response by subjecting them to a restraint stress, a stimulus known to induce stress-induced hyperthermia^{1,19}. The body temperature of EP₃-deficient mice rose like that of wild-type mice, indicating that they may maintain the efferent pathway to fever generation intact.

PGE₂ was proposed to be a central mediator of fever more than

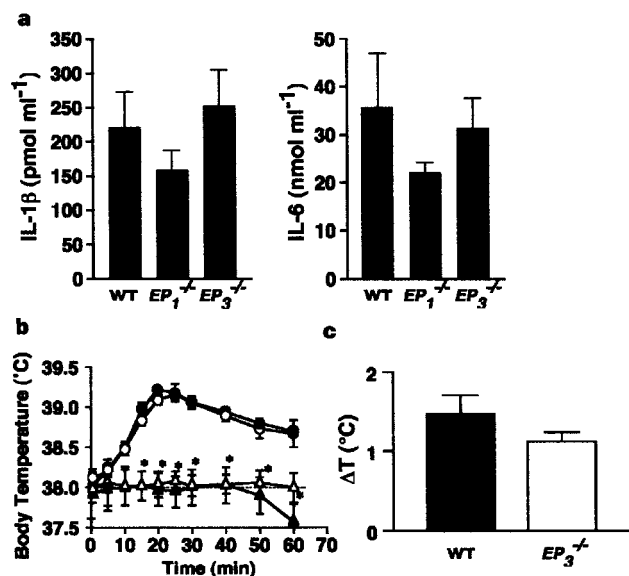


Figure 4 Responses to LPS in N4 mice lacking the EP₁ or EP₃ receptor. **a**, LPS-stimulated cytokine production by peritoneal macrophages from wild-type, EP₁^{-/-} and EP₃^{-/-} mice. IL-1β (left) and IL-6 (right) production is shown ($n = 5-10$). **b**, LPS-induced fever. LPS or vehicle was injected i.v. Open circles, wild-type mice with LPS ($n = 7$); filled circles, EP₁^{-/-} mice with LPS ($n = 5$); filled triangles, EP₃^{-/-} mice with LPS ($n = 4$); open triangles, wild-type mice with vehicle ($n = 5$). Asterisk indicates $P < 0.01$ for EP₃^{-/-} mice versus wild-type mice injected with LPS. **c**, Restraint-induced hyperthermia. Mice were subjected to restraint stress and the body temperature rise in 30 min is shown ($n = 6$, each). There was no significant difference between the wild-type and EP₃^{-/-} mice.

two decades ago³, but this role has been debated ever since^{1,4-7}. One argument against such a role was based on the inability of a PGE₂ antagonist, SC-19920, to inhibit pyrogen-induced fever⁵. Prostanoids other than PGE₂ were therefore suggested to be the fever mediator⁶, and an action other than inhibition of prostaglandin synthesis was proposed to be the functional mechanism of antipyretic drugs^{1,11,20}. We have now shown that the PGE₂-receptor subtype EP₃ is essential for the febrile response in mice. Given the binding specificity of this receptor and the facts that the potency of PGE₂ is at least 100 times that of other prostanoids and that SC-19920 shows low affinity for the EP₃ receptor²¹, our results indicate that PGE₂ is an essential mediator of fever induced by both exogenous and endogenous pyrogens. Our data thus answer the longstanding question about the role of PGE₂.

Blood-borne pyrogens may act on cells of the organum vasculosum lamina terminalis (OVLT), an area of the brain lacking the blood-brain barrier, to induce prostaglandin synthesis⁷. EP₃ receptor mRNA is abundant in neurons surrounding the OVLT²². Furthermore, PGE₂-induced fever is phylogenetically conserved from lower invertebrates, such as the leech and cricket, to mammals¹. Molecular phylogenetic analysis²³ indicates that the most primitive prostanoid receptor is a PGE₂ receptor linked to cyclic AMP metabolism that is similar to the EP₃ receptor. Fever is one of the acute-phase reactions exhibited by animals in response to injurious stimuli. Prostaglandins have also been suggested to be important in eliciting other acute-phase reactions such as the hypothalamus/adrenal cortex response²⁴. Whether the EP₃ receptor also contributes to these responses remains to be determined. □

Methods

Gene targeting. Mice lacking the EP₁ or EP₃ receptors were generated nearly as described^{25,26}. The targeting vector for the EP₁ receptor gene was made by inserting the neomycin-resistance gene (*neo*) (pMC1-*neo*; Stratagene) into the *FspI* site of exon 2, which is located immediately before the sequence encoding

the first transmembrane domain. The herpes simplex virus thymidine kinase gene (*TK*) was inserted upstream of *neo* (Fig. 1a). The targeting vector for the EP₃ receptor gene was constructed by replacing a 1-kilobase fragment containing almost all of exon 1, which encodes the region of the receptor from the amino terminus to the sixth transmembrane domain, with the β-galactosidase gene (*LacZ*) and *neo*, with *TK* inserted downstream (Fig. 1b). These targeting vectors were linearized and introduced into E14-1 embryonic stem cells by electroporation. Clones resistant to G418 and ganciclovir were isolated and screened by the polymerase chain reaction for homologous recombination, which was then confirmed by Southern blot hybridization. The generation of chimaeras and mutant mice has been described²⁷. For northern blot analysis, polyadenylated RNA (3 μg) from the kidneys of wild-type, EP₁^{-/-} and EP₃^{-/-} mutant mice was hybridized with corresponding complementary RNA probes.

Ligand-binding assay. Kidneys were cut along the longitudinal plane, and the medulla was isolated. The tissue was homogenized and a crude membrane fraction was prepared as described²⁶. Membranes (150 μg protein) were incubated at 30 °C with 4 nM [³H]PGE₂ in the presence or absence of 0.1 μM M&B28767, an EP₃ receptor agonist, and EP₃ receptor binding was determined.

Body-temperature measurements. Rectal temperature was measured with a digital thermometer (TD-300; Shibaura Electronics) by inserting the probe into the rectum 3–4 cm up from the anus at the indicated times. PGE₂ (1 nmol) in 5 μl physiological saline containing 5% ethanol was injected i.c.v. (0.5 mm posterior to the bregma and 0.7 mm lateral from the midline) with a 27-gauge needle with a stopper at 3 mm from the tip. LPS from *Escherichia coli* 026:B6 (Sigma) was dissolved in physiological saline at a concentration of 30 mg ml⁻¹ and was injected into the tail vein at a dosage of 10 mg per kg body weight. Murine recombinant IL-1β (Pepro Tech) was dissolved in phosphate-buffered saline and was injected i.v. (100 ng in 50 μl) or i.c.v. (10 ng in 5 μl). Mice were anaesthetized with diethyl ether before i.c.v. injection and maintained awake in a mouse holder during temperature measurement. For i.v. injection, mice were maintained in a holder for 2–3 h until the rectal temperature stabilized below 38 °C, to avoid the initial rise in temperature induced by restriction. For restraint-induced hyperthermia¹⁹, mice housed individually overnight were lifted gently and placed in a plastic holder. Rectal temperature was measured at 0 and 30 min. In all experiments, room temperature was maintained between 25 °C and 30 °C.

Assay of cytokine production. Peritoneal cells were collected from unstimulated mice by lavage with 5 ml RPMI 1640 medium containing heparin (2 units ml⁻¹). The cells were separated by centrifugation at 250 g for 5 min at 4 °C and suspended at a density of 1×10^6 ml⁻¹ in RPMI 1640 supplemented with 1% fetal bovine serum. Portions of the cell suspension (0.5 ml) were transferred to 24-well culture dishes and the cells were cultured in a CO₂ incubator at 37 °C for 3 h. Non-adherent cells were removed by washing the well twice with culture medium. Adherent macrophages were then cultured for 24 h in the absence or presence of LPS (1 μg ml⁻¹). The amounts of IL-1β and IL-6 released into the culture medium were determined by enzyme-linked immunosorbent assay (Endogen).

Data analysis. Data are expressed as means ± s.e.m. The significance of differences between data was evaluated by Student's *t*-test. A *P* value of <0.05 was considered statistically significant.

Received 7 April; accepted 9 July 1998.

1. Kluger, M. J. Fever: role of pyrogens and cryogens. *Physiol. Rev.* **71**, 93–127 (1991).
2. Vane, J. R. Inhibition of prostaglandin synthesis as a mechanism of action for aspirin-like drugs. *Nature New Biol.* **231**, 232–235 (1971).
3. Milton, A. S. & Wendlandt, S. A possible role for prostaglandin E₁ as a modulator for temperature regulation in the central nervous system of the cat. *J. Physiol. (Lond.)* **207**, 76P–77P (1970).
4. Stitt, J. T. Prostaglandin E as the neural mediator of the febrile response. *Yale J. Biol. Med.* **59**, 137–149 (1986).
5. Mitchell, D. et al. Is prostaglandin E the neural mediator of the febrile response? The case against a proven obligatory role. *Yale J. Biol. Med.* **59**, 159–168 (1986).
6. Milton, A. S. Thermoregulatory actions of eicosanoids in the central nervous system with particular regard to the pathogenesis of fever. *Ann. NY Acad. Sci.* **559**, 392–410 (1989).
7. Saper, C. B. & Breder, C. D. The neurologic basis of fever. *N. Engl. J. Med.* **330**, 1880–1886 (1994).
8. Ushikubi, F., Hirata, M. & Narumiya, S. Molecular biology of prostanoid receptors: an overview. *J. Lipid Mediat.* **12**, 343–359 (1995).
9. Segi, E. et al. Patent ductus arteriosus and neonatal death in prostaglandin receptor EP₄-deficient mice. *Biochem. Biophys. Res. Commun.* **246**, 7–12 (1998).
10. Nguyen, M. T. et al. The prostaglandin receptor EP₄ triggers remodelling of the cardiovascular system at birth. *Nature* **390**, 78–81 (1997).
11. Cranston, W. I. et al. Evidence that brain prostaglandin synthesis is not essential in fever. *J. Physiol. (Lond.)* **259**, 239–249 (1976).
12. Zheng, H. et al. Resistance to fever induction and impaired acute-phase response in interleukin-1β-deficient mice. *Immunity* **3**, 9–19 (1995).

13. Kozak, W. *et al.* Thermal and behavioral effects of lipopolysaccharide and influenza in interleukin-1 β -deficient mice. *Am. J. Physiol.* **269**, R969–R977 (1995).
14. Davatelis, G. *et al.* Macrophage inflammatory protein-1: a prostaglandin-independent endogenous pyrogen. *Science* **243**, 1066–1068 (1989).
15. Zamproni, A. R. *et al.* Interleukin-8 induces fever by a prostaglandin-independent mechanism. *Am. J. Physiol.* **266**, R1670–R1674 (1994).
16. Bishai, I., Dinarello, C. A. & Cocceani, F. Prostaglandin formation in feline cerebral microvessels: effect of endotoxin and interleukin-1. *Can. J. Physiol. Pharmacol.* **65**, 2225–2230 (1987).
17. Di Pauli, R. Genetics of the immune response. I. Differences in the specificity of antibodies to lipopolysaccharides among different strains of mice. *J. Immunol.* **109**, 394–400 (1972).
18. Meng, F. & Lowell, C. A. Lipopolysaccharide (LPS)-induced macrophage activation and signal transduction in the absence of Src-family kinases Hck, Fgr, and Lyn. *J. Exp. Med.* **185**, 1661–1670 (1997).
19. Morley, R. M., Conn, C. A., Kluger, M. J. & Vander, A. J. Temperature regulation in biotelemetered spontaneously hypertensive rats. *Am. J. Physiol.* **258**, R1064–R1069 (1990).
20. Alexander, S. J., Cooper, K. E. & Veale, W. L. Sodium salicylate: alternate mechanism of central antipyretic action in the rat. *Pflügers Arch.* **413**, 451–455 (1989).
21. Kiriya, M. *et al.* Ligand binding specificities of the eight types and subtypes of the mouse prostaglandin receptors expressed in Chinese hamster ovary cells. *Br. J. Pharmacol.* **122**, 217–224 (1997).
22. Sugimoto, Y. *et al.* Distribution of the messenger RNA for the prostaglandin E receptor subtype EP₃ in the mouse nervous system. *Neuroscience* **62**, 919–928 (1994).
23. Toh, H., Ichikawa, A. & Narumiya, S. Molecular evolution of receptors for eicosanoids. *FEBS Lett.* **361**, 17–21 (1995).
24. Katsura, G., Arimura, A., Koves, K. & Gottschall, P. E. Involvement of organum vasculosum of lamina terminalis and preoptic area in interleukin 1 β -induced ACTH release. *Am. J. Physiol.* **258**, E163–E171 (1990).
25. Murata, T. *et al.* Altered pain perception and inflammatory response in mice lacking prostacyclin receptor. *Nature* **388**, 678–682 (1997).
26. Sugimoto, Y. *et al.* Failure of parturition in mice lacking the prostaglandin F receptor. *Science* **277**, 681–683 (1997).
27. Kawabe, T. *et al.* The immune responses in CD40-deficient mice: impaired immunoglobulin class switching and germinal center formation. *Immunity* **1**, 167–178 (1994).

Acknowledgements. We thank Y. Kataoka and K. Ishikawa for injection of embryonic stem cells, K. Takahashi for animal breeding, and K. Okuyama and T. Arai for secretarial assistance. This work was supported by grants-in-aid from the Ministry of Education, Science, Sports and Culture of Japan and from the Ministry of Health and Welfare of Japan, and by grants from the Uehara Memorial Foundation, the Smoking Research Foundation and the Japanese Foundation on Metabolism and Diseases.

Correspondence and requests for materials should be addressed to S.N. (e-mail: snaru@mfour.med.kyoto-u.ac.jp).

Toll-like receptor-2 mediates lipopolysaccharide-induced cellular signalling

Ruey-Bing Yang, Melanie R. Mark, Alane Gray, Arthur Huang, Ming Hong Xie, Min Zhang, Audrey Goddard, William I. Wood, Austin L. Gurney & Paul J. Godowski

Department of Molecular Biology, Genentech, South San Francisco, California 94080-4990, USA

Vertebrates and invertebrates initiate a series of defence mechanisms following infection by Gram-negative bacteria by sensing the presence of lipopolysaccharide (LPS), a major component of the cell wall of the invading pathogen¹. In humans, monocytes and macrophages respond to LPS by inducing the expression of cytokines, cell-adhesion proteins, and enzymes involved in the production of small proinflammatory mediators. Under pathophysiological conditions, LPS exposure can lead to an often fatal syndrome known as septic shock². Sensitive responses of myeloid cells to LPS require a plasma protein called LPS-binding protein and the glycosylphosphatidylinositol-anchored membrane protein CD14. However, the mechanism by which the LPS signal is transduced across the plasma membrane remains unknown³. Here we show that Toll-like receptor 2 (TLR2) is a signalling receptor that is activated by LPS in a response that depends on LPS-binding protein and is enhanced by CD14. A region in the intracellular domain of TLR2 with homology to a portion of the interleukin (IL)-1 receptor that is implicated in the activation of the IL-1-receptor-associated kinase is required for this response. Our results indicate that TLR2 is a direct mediator of signalling by LPS.

In *Drosophila*, the Toll receptor participates in establishing the dorsoventral pattern in the embryo and inducing an antifungal immune response in the adult fly^{4,5}. Toll is a type I transmembrane protein containing an extracellular domain with leucine-rich

repeats and a cytoplasmic domain with sequence homology to the interleukin-1 receptor (IL-1R)^{6,7} and several plant disease-resistance proteins^{8,9}. Activation of Toll leads to induction of genes through induction of the NF- κ B pathway^{4,10,11}. In humans, five homologues of *Drosophila* Toll, designated as Toll-like receptors (TLRs), have been identified^{12,13}. A constitutively active version of one human TLR (Toll-protein homologue¹², TLR4¹³) can activate NF- κ B and induce expression of inflammatory cytokines and co-stimulatory molecules¹², suggesting that human TLRs participate in the innate immune response and signal the activation of adaptive immunity¹².

To determine whether human TLRs are involved in LPS-induced cell activation, we first investigated the expression of TLRs in immune tissues. TLR2 was found to be expressed in all lymphoid tissues examined, with the highest expression in peripheral blood leukocytes (Fig. 1a). Expression of TLR2 messenger RNA is enriched in monocytes/macrophages, the primary CD14-expressing and LPS-responsive cells. TLR2 mRNA is upregulated upon stimulation of isolated monocytes/macrophages with LPS (Fig. 1b), as has been reported for CD14 (refs 14, 15). We therefore investigated whether TLR2 could be involved in LPS-mediated cellular signalling.

We isolated a stable population and a stable clone of human embryonic kidney 293 cells that express an epitope-tagged version of TLR2 (gD-TLR2) (Fig. 2b). We tested the response of these cells to LPS prepared from *Escherichia coli* strain 055:B5, to LPS-binding protein (LBP), or to a combination of LPS and LBP, by monitoring expression of a reporter gene driven by the NF- κ B-responsive enhancer of the E-selectin gene¹⁶. Although neither LPS nor LBP treatment alone caused significant gene activation, addition of both LPS and LBP resulted in a substantial induction of reporter gene activity in cells expressing TLR2, but not in the parental 293 cells (Fig. 2a). Furthermore, LPS and LBP induced NF- κ B activity in TLR2-expressing cells (Fig. 2c) with kinetics similar to those observed in myeloid and non-myeloid cells^{17,18}. Activation of NF- κ B by LPS/LBP in 293-TLR2 cells does not require *de novo* protein synthesis, because pretreatment with cycloheximide (Fig. 2c) or actinomycin D (not shown) does not inhibit the response.

Both the membrane-bound, glycosylphosphatidylinositol (GPI)-linked CD14 (mCD14), which is present on myeloid cells, and soluble CD14 (sCD14), which is present in plasma¹⁹, enhance the

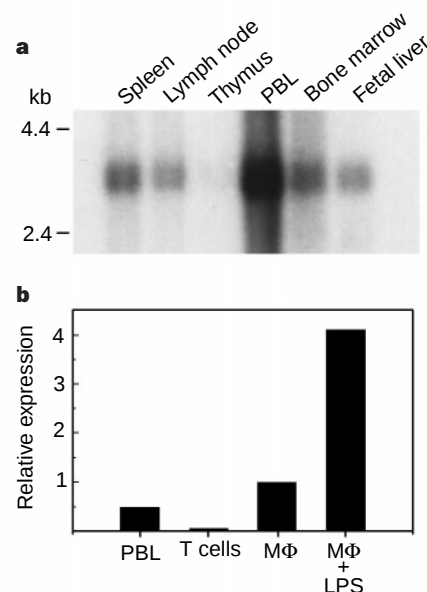


Figure 1 Expression pattern of TLR2. **a**, Northern analysis of human multiple immune tissues hybridized with a TLR2 probe. PBL, peripheral blood leukocytes. **b**, Quantitative polymerase chain reaction with reverse transcription (RT-PCR) was used to determine the relative expression of TLR2 in PBL, T cells, macrophages (MΦ), and LPS-stimulated macrophages (MΦ + LPS).

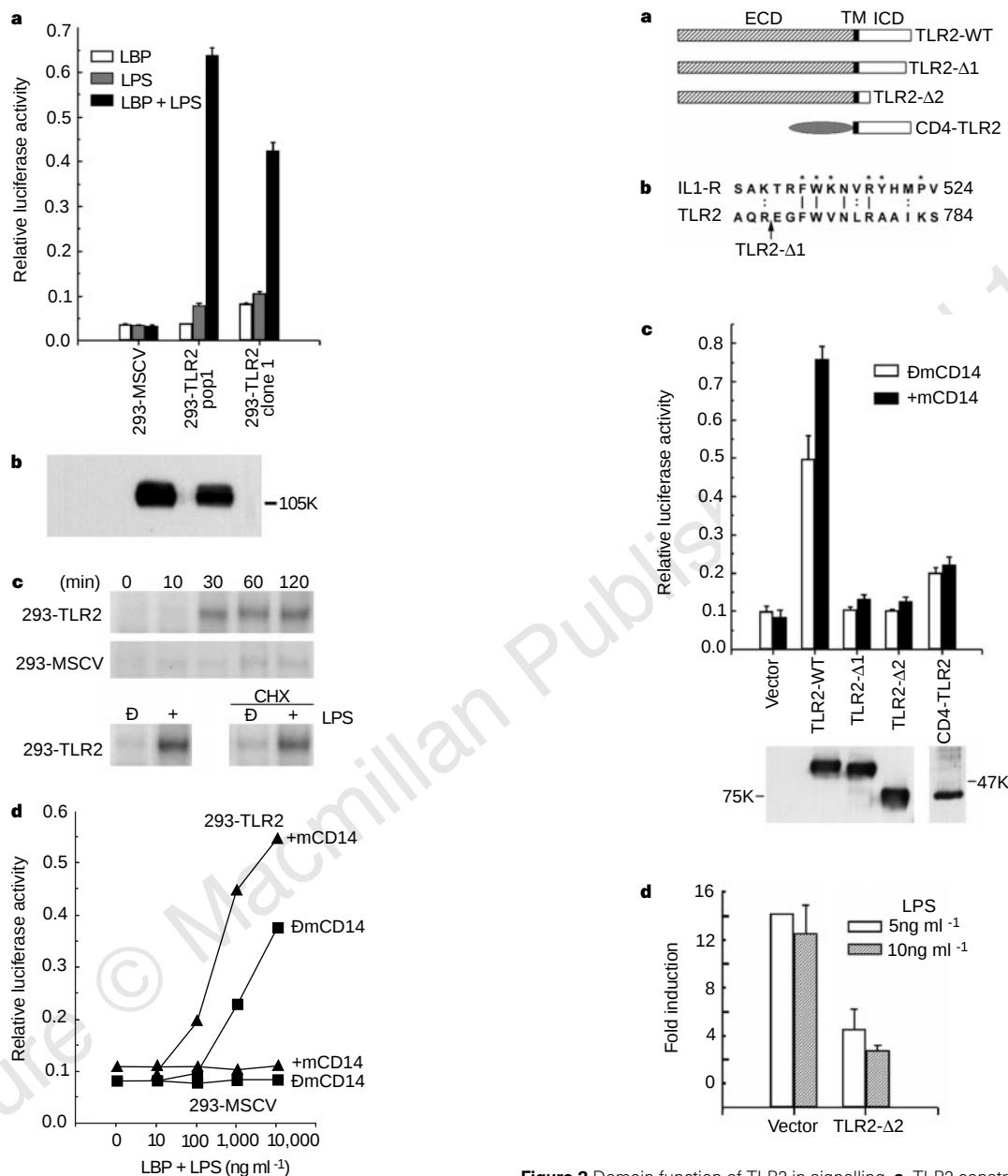


Figure 2 TLR2 mediates LPS-induced signalling. **a**, 293 cells stably expressing TLR2 acquire LPS responsiveness. We transiently transfected a population of clones expressing gD.TLR2 (293-TLR2 pop1), or a single clone expressing gD.TLR2 (293-TLR2 clone 1), or control cells (293-MSCV) with pGL3.ELAM.tk and then stimulated them with 1 μ g ml⁻¹ 055:B5 LPS for 6 h with or without LBP in serum-free medium. Activation of the ELAM enhancer was measured as described in Methods. Results were obtained from two independent experiments. No stimulation was observed using the control reporter plasmid that lacked the ELAM enhancer (data not shown). Expression of the reporter plasmid was equivalent in untreated cells or cells treated with LBP alone (data not shown). **b**, Western blot showing expression of epitope-tagged TLR2 in 293 cells. **c**, Time course of TLR2-dependent LPS-induced activation and translocation of NF- κ B. Nuclear extracts were prepared from cells treated with 055:B5 LPS (10 μ g ml⁻¹) and LBP for the indicated times (top), or cells pretreated with 1 μ M cycloheximide (CHX) for 1 h then stimulated with 1 μ g ml⁻¹ LPS for 1 h in the presence of LBP in serum-free medium (bottom). **d**, Effect of mCD14 on NF- κ B activation by TLR2. Vector control 293-MSCV (open symbols) or 293-TLR2 pop1 cells (closed symbols) were transfected with the reporter plasmid, and either a CD14 expression vector (+mCD14; triangles) or vector control (-mCD14; squares). After 24 h, transfected cells were stimulated with 055:B5 LPS for 6 h in the presence of LBP in serum-free medium. Data are representative of three independent experiments.

Figure 3 Domain function of TLR2 in signalling. **a**, TLR2 constructs. TLR2-WT is the full-length epitope-tagged form of TLR2. TLR2-Δ1 and -Δ2 represent truncations of 13 or 141 amino acids at the C terminus, respectively. CD4-TLR2 contains amino acids 1–205 of human CD4 fused to the transmembrane and intracellular domains of TLR2. ECD, extracellular domain; TM, transmembrane region; ICD, intracellular domain. **b**, C-terminal residues critical for IL-1R and TLR2 signal transduction. Residue numbers are shown to the right of each protein. Arrow indicates the position of the TLR2-Δ1 truncation. Asterisks indicate residues essential for IL-1R signalling²⁶; I, identical amino acid; colon dots, conservative changes. **c**, TLR-2 variants fail to induce NF- κ B in response to LPS and LBP. 293 cells were transiently transfected with pGL3.ELAM.tk and response to LPS and LBP. 293 cells were transiently transfected with pGL3.ELAM.tk and expression vectors encoding TLR2 or TLR2 variants, as indicated. Cells were also transfected with a CD14 expression plasmid (+mCD14) or with a control expression plasmid that lacks any cDNA insert (-mCD14). Expression of the TLR2 variants was confirmed by western blotting using either anti-gD or CD4 antibody (bottom). The luciferase assay was done as described in Methods. Data were from duplicate experiments. **d**, Inhibition of endogenous LPS responsiveness by TLR2-Δ2. LPS-responsive U373 cells, transfected with reporter, mCD14 and vector alone or TLR2-Δ2 constructs, were stimulated with LPS and either 5 or 10 ng ml⁻¹ of LPS from *E. coli* K12 (strain LCD25) for 6 h. The luciferase activity (average of duplicate wells) was determined and expressed as fold induction compared with unstimulated U373 cells transfected with reporter, mCD14 and control vector.

responsiveness of cells to LPS. We found that 293 cells express little or no CD14 on their cell surface (data not shown). However, transient transfection of 293 cells with mCD14 increased the sensitivity and magnitude of TLR2-mediated LPS responsiveness (Fig. 2d).

The intracellular domain of TLR2 is important in signalling the LPS response because variants with carboxy-terminal truncations of either 13 (TLR2- Δ 1) or 141 amino acids (TLR2- Δ 2) were defective for induction of the reporter in transiently transfected 293 cells (Fig. 3a, c). The region of the intracellular domain deleted in TLR2- Δ 1 is homologous to a region of the IL-1R intracellular domain that is required for association with the IL-1R-associated kinase IRAK¹⁶ (Fig. 3b). The extracellular domain (ECD) of TLR2 is also required for responsiveness to LPS because a TLR2 variant in which the ECD of TLR2 was replaced with a portion of the ECD of CD4 failed to respond to LPS (Fig. 3a, c).

We investigated whether TLR2- Δ 2 acted as a dominant-negative receptor to inhibit the 'endogenous' response of U373 astrocytoma cells, which express the cell-adhesion molecule ELAM and IL-6 in response to LPS²⁰. We found that the ELAM reporter construct was induced in U373 cells by low concentrations of LPS, that this response was not further modulated by expression of full-length TLR2 (data not shown), and that expression of TLR2- Δ 2 in these cells inhibited the LPS response (Fig. 3d). Expression of the ELAM reporter construct in U373 cells is also induced by the tumour-necrosis factor TNF- α , which activates NF- κ B through the TNF receptor. However, this response was not inhibited by co-expression of TLR2- Δ 2 (data not shown). Also, TLR2- Δ 2 forms heterodimers with TLR2, but not with the IL-1 receptor, in 293 cells (R.-B.Y., manuscript in preparation). Taken together, these results indicate that TLR2- Δ 2 is a specific inhibitor of LPS-mediated NF- κ B activation.

LPS is a complex glycolipid consisting of the proximal hydrophobic lipid A moiety, the distal hydrophilic O-antigen polysaccharide region, and the core oligosaccharide that joins the lipid A and O-antigen structures. There is considerable diversity in the O-antigen structures from different Gram-negative bacteria. The lipid A portion of LPS is sufficient for LPS responses, whereas treatments that remove the fatty-acid side chains from the lipid A moiety inactivate LPS. We compared the potency of LPS prepared from various Gram-negative bacteria, lipid A alone, and LPS that had been 'detoxified' by alkaline hydrolysis. LPS isolated from *Escherichia coli* serotype LCD25 was nearly two orders of magnitude more potent than the serologically distinct *Escherichia coli* 055:B5 LPS (Fig. 4a). Using 293 cells that stably express both TLR2 and mCD14, we detected activation of the ELAM reporter using as little as 400 pg ml⁻¹ LCD25 LPS (Fig. 4b). LPS prepared from *Salmonella minnesota* R595 was also a potent inducer of TLR2 activity (Fig. 4a). TLR2 was activated by lipid A alone, whereas there was no response to 'detoxified LPS', even at concentrations as high as 10 μ g ml⁻¹ (Fig. 4a, and data not shown).

We tested whether TLR2 binds LPS by using a soluble form of the TLR2 extracellular domain (TLR2-Fc) and ³H-labelled LPS *in vitro*. ³H-LCD25 LPS bound to the TLR2-Fc fusion protein but not to fusion proteins containing the ECDs of several other receptors (Fig. 4c). This binding was specifically competed by cold LCD25 LPS but not by detoxified LPS. Preliminary analysis of the binding of LPS to TLR2-Fc suggests that the *K*_d is relatively low (500–700 nM) and that the kinetics of association are slow (data not shown). Other proteins, such as LBP, may enhance binding of LPS to TLR2 *in vivo*, much like LBP acts to transfer LPS from its free, aggregated (micellar) form to CD14. Nevertheless, TLR2-Fc could be used to immunodeplete LPS, as assayed by the inhibition of LPS-induced TNF- α release from primary macrophages (Fig. 4d).

LPS treatment of macrophages leads to the expression of several inflammatory cytokines. Likewise, expression of TLR2 in 293 cells resulted in a >100 fold-induction of IL-8 mRNA in response to LPS/LBP, whereas detoxified LPS was inactive in this assay (Fig. 5).

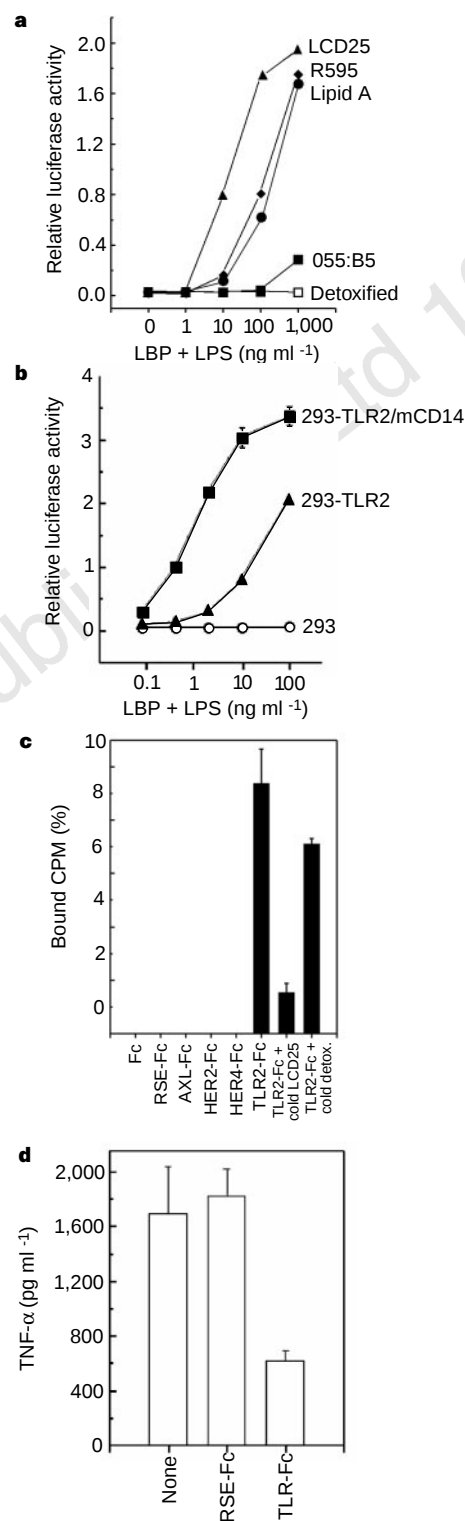


Figure 4 High potency of *E. coli* K12 LPS (LCD25) and its binding to TLR2. **a**, Dose-response curve for activation of the ELAM reporter gene by different types of LPS in 293-TLR2 cells. **b**, Activation of TLR2 by low concentrations of LPS in 293 stably expressing both TLR2 and mCD14. **c**, Specific interaction of ³H-LPS (LCD25) with the extracellular domain of TLR2. We observed specific binding to TLR2-Fc, but not to Fc alone, or fusion proteins containing the extracellular domains of tyrosine kinase receptors Rse, Axl, Her2 or Her4. Binding to TLR2-Fc was specifically competed with cold LCD25 LPS, but not with detoxified LPS. **d**, Inhibition of LPS-induced TNF- α production by TLR2-Fc. Heparinized human whole blood was exposed to LCD25 LPS preincubated with buffer alone ('None') or with RSE-Fc or TLR2-Fc fusion protein. TNF- α levels were measured by enzyme-linked immunoassay. Unstimulated whole blood released no detectable TNF- α (<15 pg ml⁻¹).

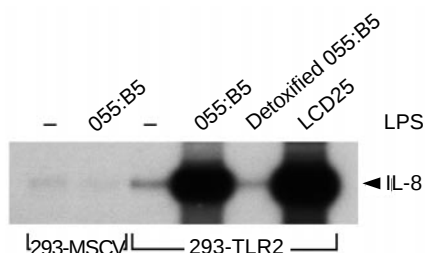


Figure 5 Activation of TLR2 by LPS in 293 cells results in induction of IL-8 mRNA expression. Control 293-MSCV and 293-TLR2 cells transiently expressing mCD14 were stimulated with LBP alone (-), or together with the indicated type of LPS ($1 \mu\text{g ml}^{-1}$) in serum-free medium for 6 h. The blot was probed for IL-8 mRNA.

These results indicate that TLR2 plays a sentinel role in the innate immune response, being the first line of defence against microbial pathogens. TLR2 participates in sensing LPS by a mechanism that is dependent on LBP and is enhanced by CD14. After LPS is detected, TLR2 transmits the information across the plasma membrane, resulting in the activation of NF- κ B and regulation of responsive genes.

Drosophila Toll and the Toll-related receptor 18 Wheeler are important in the induction of antimicrobial peptides in response to bacteria and fungi, respectively²¹. Genetic data have implicated Spätzle as a ligand for Toll in dorsoventral patterning and led to speculation that homologues of Spätzle are ligands of human TLRs^{21,22}. Our data are consistent with a model in which TLR2 participates in LPS recognition, but do not exclude the possibility that other proteins may modify the response of TLR2 to LPS. We note that although extracellular domains of TLR2 and *Drosophila* Toll both contain leucine-rich repeats, they share less than 20% amino-acid identity. Second, proteins with leucine-rich repeats are pattern-recognition receptors for various types of molecules, such as proteins, peptides and carbohydrates⁸. Third, the requirement for Spätzle in the *Drosophila* immune response is less well understood than for Toll. Unlike Toll mutants, Spätzle mutants induce normal levels of the antimicrobial peptides defensin and attacin and are only partially defective in cecropin A expression following fungal challenge⁵, and are not defective in activation of Dorsal in response to injury²³.

Our results indicate that TLR receptors act as pathogen-pattern-recognition receptors, sensing the presence of conserved molecular structures on microbes, as proposed earlier²⁴. TLR2 is a potential target in devising a treatment for septic shock. □

Methods

Reagents. ³H-labelled, unlabelled LCD25 and *S. minnesota* R595 LPS were from List Biochemicals; all other LPS were from Sigma. LBP was supplied as serum-free conditioned medium from 293 cells transfected with a human LBP expression vector. The TLR2-Fc fusion protein was produced by a baculovirus system and purified as described²⁵.

Construction of expression plasmids. A cDNA encoding human TLR2 was cloned from a human fetal lung library. The predicted amino-acid sequence matched that previously published¹³, except for a Glu-to-Asp substitution of amino acid 726. The N-terminal epitope tag version of TLR2 (gD.TLR2) was constructed by adding an *Xho*I restriction site immediately upstream of leucine at position 17 (the first amino acid of the predicted mature form of TLR2) and linking this to amino acids 1–53 of herpes simplex virus type 1 glycoprotein D as described²⁵. PCR products were sequenced and subcloned into a mammalian expression vector that contains the puromycin resistance gene. C-terminal truncation variants of gD.TLR2 were constructed by digestion of the cDNA at either a *B*spI (variant Δ 1) or *N*siI (variant Δ 2) site present in the coding sequence of the intracellular domain and at a *N*otI site present in the 3' polylinker of the expression vector followed by ligation of oligonucleotide linkers. Δ 1: 5'-TCA GCG GTA AGC-3' and 5'-GGC CGC TTA CCG C-3'; Δ 2: 5'-TAA GCT TAA GC-3' and 5'-GGC CGC TTA AGC TTA TGC A-3'. The

CD4/TLR2 chimera was constructed by PCR and contained amino acids 1–205 (the signal peptide and two immunoglobulin-like domains) of human CD4 fused to amino acids 588–784 (the transmembrane and intracellular domain) of human TLR2 with a linker-encoded valine at the junction of the CD4 and TLR2 sequences. The pGL3.ELAM.tk reporter plasmid contained the sequence 5'-GGT ACC TTC TGA CAT CAT TGT AAT TTT AAG CAT CGT GGA TAT TCC CGG GAA AGT TTT TGG ATG CCATTG GGG ATT TCC TCT TTA GAT CTG GCG CGG TCC CAG GTC CAC TTC GCA TAT TAA GGT GAC GCG TGT GGC CTC GAA CAC CGA GCG ACC CTG CAG CGA CCC GCA AGC TT-3' inserted between the *S*acI and *H*indIII sites of the luciferase reporter plasmid pGL3 (Promega). The C-terminal epitope tag version of LBP (LBP-Flag) was constructed by PCR through the addition of an *A*scI site in place of the native stop codon and the subcloning of this fragment into pRK5-Flag, resulting in the C-terminal addition of amino acids GRADYKDDDDK.

Stable cell lines/pools. 293 human embryonic kidney cells were grown in LGDMEM/HAM's F12 (50:50) medium supplemented with 10% FBS, 2 mM glutamine, and penicillin/streptomycin. The vector control line, 293-MSCV, was generated by stable transfection of 293 cells with the empty expression cassette. For stable expression of gD.TLR2, cells were transfected with the gD.TLR2 expression vector and selected for puromycin resistance at a final concentration of $1 \mu\text{g ml}^{-1}$. A stable pool of cells (293-TLR2 pop1) was isolated by FACS using an antibody raised against the gD tag. Both the pool and the single-cell clone (293-TLR2 clone1) were characterized by FACS and western blot analyses as described²⁵. A stable cell line expressing both CD14 and TLR2 was constructed by transfecting 293-TLR2 pop 1 with a CD14 expression vector containing a neomycin-selectable marker. A clone resistant to both neomycin and puromycin was isolated; expression of both gD.TLR2 and CD14 was verified by FACS and western blot analyses.

Luciferase reporter assay and electrophoretic mobility shift assay (EMSA). 293 parental or stable cells (2×10^5 cell per well) were seeded into six-well plates, and transfected on the following day with the expression plasmids, together with 0.5 μg of the luciferase reporter plasmid pGL3-ELAM.tk and 0.05 μg of the *Renilla* luciferase reporter vector as an internal control. After 24 h, cells were treated with either LPS, LBP or both LPS and LBP for 6 h, and the reporter gene activity was measured. Data are expressed as relative luciferase activity by dividing firefly luciferase activity with that of *Renilla* luciferase. For EMSA, nuclear extracts were prepared and used in a DNA-binding reaction with a 5'-[³²P]-radiolabelled oligonucleotide sequence containing a consensus NF- κ B binding site (Santa Cruz Biotechnology, sc-2511). The identity of NF- κ B in the complex was confirmed by supershift with antibodies against NF- κ B (data not shown). U373 cells were grown in 12-well plates (8×10^4 cell per well) as described²⁰. Cells were transfected with 0.1 μg of reporter, 0.01 μg of *Renilla* control, 0.05 μg of mCD14 expression vector and 0.5 μg of either control or TLR2- Δ 2 expression plasmid using DOSPER reagent (Boehringer-Mannheim). Cells were treated with LPS and extracts prepared as described above.

RNA expression. The tissue northern blot was purchased from Clontech and hybridized with a probe encompassing the extracellular domain of TLR2. Polyadenylated mRNA was isolated from 293 cells or 293-TLR2 cells and northern blots were probed with a human IL-8 cDNA fragment. TLR2 expression was determined using quantitative PCR using real time "Taqman" technology and analysed on a Model 770 sequence detector (Applied Biosystems) essentially as described²⁶. Forward and reverse primers, 5'-GCG GGA AGG ATT TTG GGT AA-3' and 5'-TGA GAG CTG CGA TAA AGT CCT AGG TTC CCATAT-3', labelled on the 5' nucleotide with the reporter dye FAM and on the 3' nucleotide with the quenching dye TAMRA. Macrophage/monocytes were treated for 16 h with $1 \mu\text{g ml}^{-1}$ LPS.

Receptor-binding assay. To assay binding, 20 ng ³H-LPS was mixed with 1 μg TLR2-Fc in 100 μl binding buffer (150 mM NaCl, 20 mM HEPES, 0.03% BSA) containing 15 μl protein A-Sepharose. After 3 h incubation at room temperature, protein A-Sepharose samples were washed twice with cold PBS/0.1% NP-40 and resuspended in binding buffer containing 1% SDS and 25 mM EDTA, and the amount of bound ³H-LPS was quantified by scintillation counting.

Received 10 June; accepted 12 August 1998.

1. Rietschel, E. T. & Brade, H. Bacterial endotoxins. *Sci. Am.* **267**, 54–61 (1992).
2. Parillo, J. E. Pathogenetic mechanisms of septic shock. *N. Engl. J. Med.* **328**, 1471–1477 (1993).

- Ulevitch, R. J. & Tobias, P. S. Receptor-dependent mechanisms of cell stimulation by bacterial endotoxin. *Annu. Rev. Immunol.* **13**, 437–457 (1995).
- Belvin, M. P. & Anderson, K. V. A conserved signaling pathway: the *Drosophila* toll-dorsal pathway. *Annu. Rev. Cell Dev. Biol.* **12**, 393–416 (1996).
- Lemaître, B., Nicolas, E., Michaut, L., Reichhart, J. M. & Hoffmann, J. A. The dorsoventral regulatory gene casette *spätzle/Toll/cactus* controls the potent antifungal response in *Drosophila* adults. *Cell* **86**, 973–983 (1996).
- Gay, N. J. & Keith, E. J. *Drosophila* Toll and IL-1 receptor. *Nature* **351**, 355–356 (1991).
- Heguy, A., Baldari, C. T., Macchia, G., Telford, J. L. & Melli, M. Amino acids conserved in interleukin-1 receptors (IL-1Rs) and the *Drosophila* toll protein are essential for IL-1R signal transduction. *J. Biol. Chem.* **267**, 2605–2609 (1992).
- Dangl, J. & Holub, E. La dolce vita: a molecular feast in plant–pathogen interactions. *Cell* **91**, 17–24 (1997).
- Hammond-Kosack, K. E. & Jones, J. D. G. Plant disease resistance genes. *Annu. Rev. Plant Physiol. plant Mol. Biol.* **48**, 575–607 (1997).
- Kidd, S. Characterization of the *Drosophila* cactus locus and analysis of interactions between cactus and dorsal proteins. *Cell* **71**, 623–635 (1992).
- Baeuerle, P. A. & Henkel, T. Function and activation of NF- κ B in the immune system. *Annu. Rev. Immunol.* **12**, 141–179 (1994).
- Medzhitov, R., Preston-Hurlburt, P. & Janeway, C. A. Jr A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity. *Nature* **388**, 394–397 (1997).
- Rock, F. L., Hardiman, G., Timans, J. C., Kastelein, R. A. & Bazan, J. F. A family of human receptors structurally related to *Drosophila* Toll. *Proc. Natl Acad. Sci. USA* **95**, 588–593 (1998).
- Matsuura, K. *et al.* Upregulation of mouse CD14 expression in Kupffer cells by lipopolysaccharide. *J. Exp. Med.* **179**, 1671–1676 (1994).
- Marchant, A., Duchow, J., Delville, J. P. & Goldman, M. Lipopolysaccharide induces up-regulation of CD14 molecule on monocytes in human whole blood. *Eur. J. Immunol.* **22**, 1663–1665 (1992).
- Croston, G. E., Cao, Z. & Goeddel, D. V. NF- κ B activation by interleukin-1 (IL-1) requires an IL-1 receptor-associated protein kinase activity. *J. Biol. Chem.* **270**, 16514–16517 (1995).
- Delude, R. L. *et al.* CD14-mediated translocation of nuclear factor- κ B induced by lipopolysaccharide does not require tyrosine kinase activity. *J. Biol. Chem.* **269**, 22253–22260 (1994).
- Lee, J. D. *et al.* Glycosyl-phosphatidylinositol-anchored or integral membrane forms of CD14 mediate identical cellular responses to endotoxin. *Proc. Natl Acad. Sci. USA* **90**, 9930–9934 (1993).
- Bazil, V. *et al.* Biochemical characterization of a soluble form of the 53-kDa monocyte surface antigen. *Eur. J. Immunol.* **16**, 1583–1589 (1986).
- Frey, E. A. *et al.* Soluble CD14 participates in the response of cells to lipopolysaccharide. *J. Exp. Med.* **176**, 1665–1671 (1992).
- Medzhitov, R. & Janeway, C. A. Jr Self-defense: the fruit fly style. *Proc. Natl Acad. Sci. USA* **95**, 429–430 (1998).
- Lemaître, B., Reichhart, J. M. & Hoffmann, J. A. *Drosophila* host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proc. Natl Acad. Sci. USA* **94**, 14614–14619 (1997).
- Lemaître, B. *et al.* Functional analysis and regulation of nuclear import of dorsal during the immune response in *Drosophila*. *EMBO J.* **14**, 536–545 (1995).
- Medzhitov, R. & Janeway, C. A. Jr Innate immunity: the virtues of a nonclonal system of recognition. *Cell* **91**, 295–298 (1997).
- Mark, M. R. *et al.* A novel receptor-type tyrosine kinase with homology to Axl/Ufo, is expressed at high levels in the brain. *J. Biol. Chem.* **269**, 10720–10728 (1994).
- Luoh, S. M. *et al.* Cloning and characterization of a human leptin receptor using a biologically active leptin immunoadhesin. *J. Mol. Endocrinol.* **18**, 77–85 (1997).

Acknowledgements. We thank our colleagues at Genentech, in particular members of the SPDI team, for helpful and invigorating discussions, and the oligonucleotide synthesis group for the preparation of synthetic DNA.

Correspondence and requests for materials should be addressed to P.J.G. (e-mail: ski@gene.com).

Amino-acid transport by heterodimers of 4F2hc/CD98 and members of a permease family

Luca Mastroberardino^{*,†}, Benjamin Spindler^{*,†}, Rahel Pfeiffer^{*}, Patrick J. Skelly[‡], Jan Löffing[§], Charles B. Shoemaker[‡] & François Verrey^{*}

*Institutes of * Physiology and § Anatomy, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland*

‡ Department of Immunology and Infectious Diseases, Harvard School of Public Health, 665 Huntington Avenue, Boston, Massachusetts 02115, USA

† These authors contributed equally to this work

Amino-acid transport across cellular plasma membranes depends on several parallel-functioning (co-)transporters and exchangers¹. The widespread transport system L accounts for a sodium-independent exchange of large, neutral amino acids, whereas the system y⁺L exchanges positively charged amino acids and/or neutral amino acids together with sodium^{2,3}. The molecular nature of these transporters remains unknown, although expres-

sion of the human cell-surface glycoprotein 4F2 heavy chain (h4F2hc; CD98 in the mouse)^{4,5} is known to induce low levels of L- and/or y⁺L-type transport^{6–9}. This glycoprotein is found in activated lymphocytes, together with an uncharacterized, disulphide-linked lipophilic light chain with an apparent relative molecular mass of 40,000 (*M_r* 40K)^{10,11}. Here we identify the permease-related protein E16 (ref. 12) as the first light chain of h4F2hc and show that the resulting heterodimeric complex mediates L-type amino-acid transport. The homologous protein from *Schistosoma mansoni*, SPRM1, also associates covalently with coexpressed h4F2hc glycoprotein, although it induces amino-acid transport of different substrate specificity. The coexpression of h4F2hc is required for surface expression of these permease-related light chains, which belong to a new family of amino-acid transporters that form heterodimers with cell-surface glycoproteins.

While searching for early aldosterone-regulated gene products, we cloned a complementary DNA (ASUR4) from A6 kidney cells. This cDNA encodes a highly lipophilic, permease-related protein that migrated in SDS–polyacrylamide gel electrophoresis (PAGE) as a band of ~40K (theoretical *M_r* 55.5K; *in vitro* translation not shown) (EMBL accession number Y12716)¹³. The protein ASUR4 shows 26% identity with the yeast methionine permease MUP-1 (ref. 14) and 40% identity with the membrane protein SPRM1 of the platyhelminth *S. mansoni* (GenBank accession number L25068; P.J.S. and C.B.S., manuscript in preparation). The highest degree of identity found in the database is with the human E16 protein, a partial cDNA of which had originally been cloned because it is induced in activated lymphocytes¹². The predicted amino-acid sequence of the full-length coding cDNA of E16 (P.J.S., B.M. Rosenthal and C.B.S., manuscript in preparation; GenBank accession number AF077866) indicates that it is the human homologue of ASUR4 (79% identity; sequence comparison is available as Supplementary information). The hydrophathy and TMpred plots of ASUR4, E16 and SPRM1 are very similar and support a 12-transmembrane-domain topology (Fig. 1) (ISREC TMpred server)¹⁵.

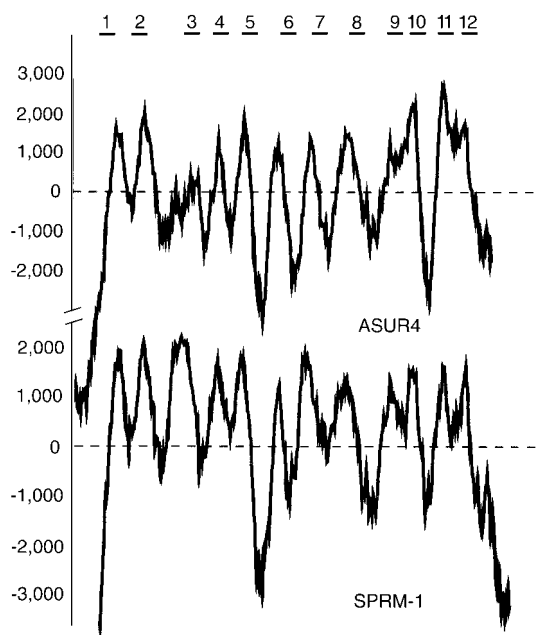


Figure 1 Transmembrane topology prediction for permease-related putative light chains. The TMpred plots indicate the probability of membrane-spanning regions for ASUR4 and SPRM1 (ISREC TMpred server). The putative transmembrane domains are indicated on the top as straight lines. The preferred orientation places the amino and carboxy termini inside the cell. The plot for E16 (not shown) can be superimposed over that of ASUR4.

Because none of these three permease-related polypeptides induced amino-acid transport when expressed in *Xenopus laevis* oocytes, we tested the hypothesis that they are light chains of 4F2hc-type glycoproteins, and hence need to form heterodimers to function as amino-acid transporters at the cell surface. As for E16, expression of h4F2hc (and of its murine homologue, CD98) and L-type amino-acid transport are enhanced in activated lymphocytes⁹. Association of h4F2hc with its uncharacterized light chain^{10,11} may be required to allow amino-acid transport¹⁶.

We first showed that any of the three lipophilic putative light chains, when coexpressed with h4F2hc for various times (1–4 days), induced amino-acid transport in *Xenopus* oocytes on a scale that was much greater than that produced by h4F2hc alone. We injected increasing amounts of putative light-chain cRNA together with a constant amount of h4F2hc cRNA. The results show that the level of amino-acid transport was proportional to the quantity of putative light-chain cRNA injected, up to a threshold that was set by the quantity of injected h4F2hc cRNA (data not shown). This shows

that h4F2hc can be limiting to the expression of functional transporter. In subsequent experiments, we injected equal amounts of heavy-chain and putative light-chain cRNA and measured amino-acid transport after 24 h (the 24-h latency period was chosen because, for a period after injection, oocytes injected with h4F2hc alone showed amino-acid-transport levels indistinguishable from those of non-injected oocytes).

Coexpressed ASUR4 and h4F2hc transported leucine at all tested concentrations (from 1 μ M to 10 mM), independently of the presence of Na⁺ (Fig. 2a and data not shown), as reported for system L^{1,2}. This transport was inhibited by the system-L-specific substrate 2-(–)-endoamino-bicycloheptane-2-carboxylic acid (BCH) but not by the system-A-specific substrate α -(methylamino)isobutyric acid (MeAIB) (Fig. 2a). As expected for system L, the efflux of leucine was dependent on the presence of leucine in the buffer, showing that amino-acid exchange occurred (Fig. 2b). The transporter accepted a broad range of large, neutral amino acids (Figs 2c and 3), although with differing apparent K_m values ranging from 32 μ M for leucine to 2.2 mM for glutamine. Maximal transport rates were comparable for all substrates tested. Transport of glutamate, arginine and lysine was not significantly increased compared with oocytes injected with h4F2hc alone at the substrate concentration used (100 μ M).

A comparison of the substrate specificity and Na⁺ dependence of the three transport systems (ASUR4, E16 or SPRM1 + h4F2hc) is shown in Fig. 3. E16 shows a pattern very similar to that of ASUR4 when coexpressed with h4F2hc and, therefore, corresponds functionally to human system L. We assayed leucine uptake at different concentrations (1, 10, 100 and 1,000 μ M) in the presence or absence of Na⁺ for all three heterodimeric transporters (data not shown). These experiments confirm the similarity of leucine transport induced by either ASUR4 or E16 coexpressed with h4F2hc; transport in both cases was not significantly affected by the presence of Na⁺ and showed similar K_m values (39 ± 16 and 80 ± 23 μ M, $n = 6$, respectively).

There were three major differences between the system-L-type transport induced by ASUR4 and E16 and that induced by the *S. mansoni* protein SPRM1 when coexpressed with h4F2hc. First, SPRM1 transported cationic amino acids in a manner similar to transport by system γ^+ or γ^+L . Second, it showed a partial Na⁺

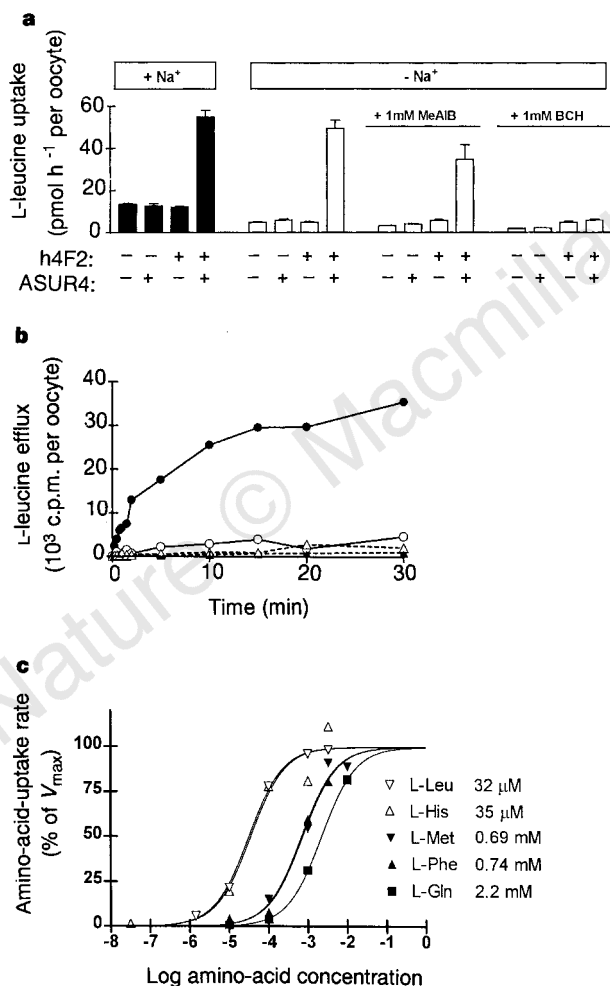


Figure 2 Amino-acid uptake and efflux in *Xenopus* oocytes expressing h4F2hc and/or the putative light chain ASUR4. **a**, L-leucine (10 μ M) uptake is independent of the presence of Na⁺ and inhibited by the system-L-specific substrate BCH but not the system-A-specific substrate MeAIB. **b**, L-leucine efflux was measured from oocytes injected with h4F2hc alone (dashed line) or h4F2hc + ASUR4 (solid line) cRNA. In the presence of 1 mM cold L-leucine (filled symbols), but not in its absence (empty symbols), an efflux of radioactive L-leucine was observed from injected oocytes. **c**, Concentration dependence of the uptake of five different amino acids by h4F2hc + ASUR4. Sigmoidal curves corresponding to Michaelis-Menten kinetics were fitted to the experimental points to determine the apparent K_m values.

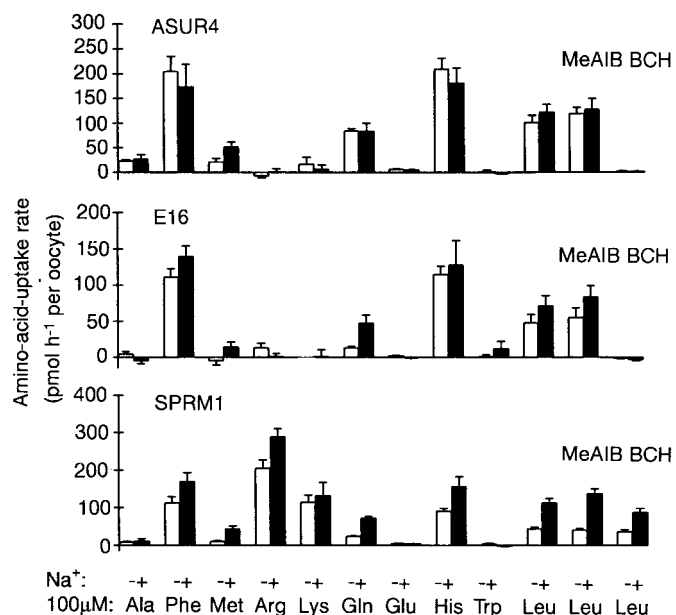


Figure 3 Specificity and Na⁺ dependence of amino-acid uptake vary when h4F2hc is expressed together with different putative light chains (ASUR4, E16 or SPRM1). Uptakes were for 4 min; MeAIB and BCH were added at a final concentration of 10 mM.

dependence for the transport of neutral amino acids, in contrast to system L which is Na⁺-independent, but less than would be expected for system y⁺L. Its apparent K_m for the transport of leucine (200–500 μM) was also higher than that for the two L-type transporters. Third, it was not inhibited by the system-L-specific substrate BCH. The mixed properties of this transport mediated by the platyhelminth SPRM1 when coexpressed with h4F2hc (which resembles systems y⁺L and L) do not fit the definition of any mammalian amino-acid-transport system. However, these properties show that the characteristics of the transport can vary considerably depending on which putative light chain is expressed with h4F2hc.

To test biochemically whether the permease-like proteins associate covalently with 4F2hc, as expected for light chains, we performed immunoprecipitation experiments. h4F2hc expressed alone

in *Xenopus* oocytes produced a band of ~80K on SDS–PAGE upon immunoprecipitation with anti-h4F2hc antibody (Fig. 4, lane 2). But coexpression of h4F2hc together with ASUR4 led to the appearance of an extra band at ~40K, which corresponds to coprecipitated ASUR4 (Fig. 4, lane 3)^{17,18}. When unreduced immunoprecipitate was run on SDS–PAGE, h4F2hc and ASUR4 remained attached, indicating that they are linked by a disulphide bond (Fig. 4, lane 8). A similar phenomenon was observed when either E16 or SPRM1 was expressed together with h4F2hc and immunoprecipitated with anti-h4F2hc antibody (Fig. 4, lanes 4, 5, 9 and 10), which confirms that these permease-related proteins are light chains of 4F2hc-type glycoproteins.

A model for the membrane topology of an amino-acid-transporter heterodimer is shown in Fig. 5. It is based on the 12-transmembrane-domain topology proposed by the TMpred analysis (Fig. 1) and on the fact that the extracellular cysteine residue involved in the disulphide bridge with the heavy chain is located in the loop between predicted transmembrane helices 3 and 4 (R.P., B.S., L.M. and E.V., unpublished results).

We propose the name hAmAT-L for the human heterodimeric amino-acid transporter type L described here, and the name hAmAT-L-lc for the corresponding permease-like multitransmembrane subunit (E16).

To determine whether association of the AmAT-type light chains with h4F2hc is required for cell-surface expression of the light chains in *Xenopus* oocytes, double immunofluorescence localization experiments were done using antibodies against h4F2hc and SPRM1. h4F2hc is surface-localized when expressed with or without

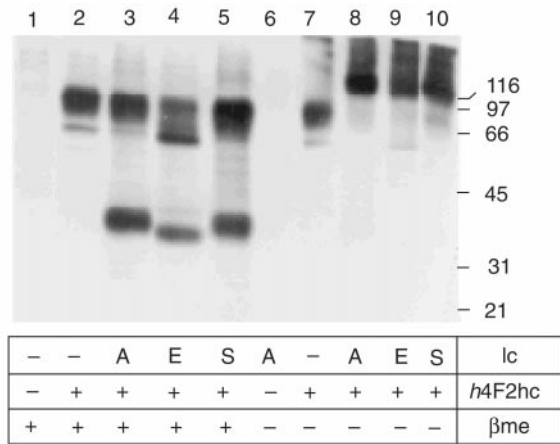


Figure 4 Association of h4F2hc with different light chains: immunoprecipitation of h4F2hc ± light chains (lc) expressed in *Xenopus* oocytes with anti-h4F2hc antibody followed by SDS–PAGE and fluorography (A, ASUR4; E, E16; S, SPRM1). Immunoprecipitation of h4F2hc co-expressed with light chain shows additional ~40K bands corresponding to the co-immunoprecipitated light chains when the proteins were resolved on the gel after reduction with β-mercaptoethanol (βme) (lanes 3–5). Under unreducing conditions, h4F2hc and the light chains appear as ~120K apparent heterodimers (lanes 8–10).

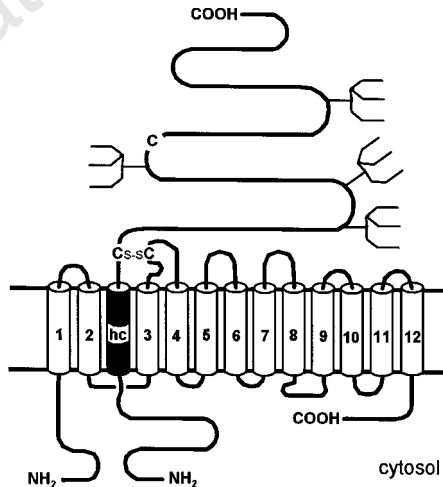


Figure 5 Membrane-topology model of a heterodimeric amino-acid transporter of the AmAT family. The label hc indicates the single transmembrane domain of the glycoprotein heavy chain (h4F2hc). Potential N-glycosylation sites are indicated by forks. The putative transmembrane domains of the lipophilic light chain (ASUR4, E16 or SPRM1) are numbered 1–12. The putative cysteine residues involved in disulphide linkage are indicated.

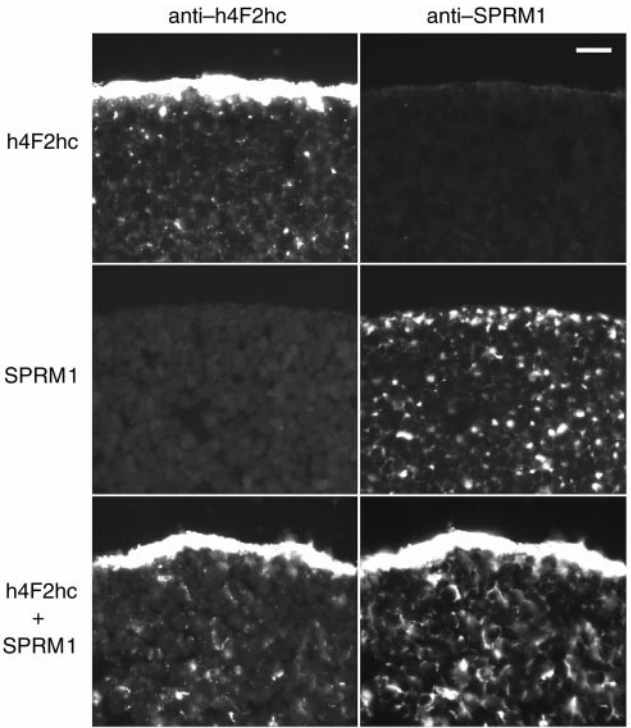


Figure 6 Double immunofluorescence detection of h4F2hc and the light-chain SPRM1 in oocyte sections. Injected cRNAs are indicated at the left side of each panel pair (same section), and the detecting antibody is indicated at the top of each lane. Most of the h4F2hc expressed alone in oocytes is localized at the cell surface, although some is also present in a reticular and punctuate intracellular compartment (top left panel). In contrast, SPRM1 expressed alone is localized only in reticular and punctuate structures throughout the cell, and not at the cell surface (middle right panel). On co-expression of the two subunits, SPRM1 localizes together with h4F2hc mainly at the cell surface and also in intracellular reticular and punctuated structures (lower panels). Bar, 5 μm.

SPRM1, whereas SPRM1 is surface-localized in oocytes only when expressed with h4F2hc (Fig. 6). These results thus support the hypothesis that the 4F2hc type II glycoprotein is necessary for the maturation, transport and/or surface residence of the AmAT-Ic subunits. This phenomenon is analogous to that seen with the Na^+/K^+ -ATPase, in which the β -subunit glycoprotein is required for surface expression of its catalytic multitransmembrane domain α -subunit(s)¹⁹.

The heterologous expression of the type II transmembrane heavy chain h4F2hc alone facilitates amino-acid transport not only through the L-type, but also through the γ^+ L-type, system, depending on the cellular system^{6,7,9,16}. This indicates that in addition to hAmAT-L-Ic (E16), at least one other related light chain can associate with h4F2hc to produce a heterodimeric transporter with γ^+ L-transport characteristics (hAmAT- γ^+ L). It will be interesting to test this hypothesis, which has implications for the regulation of amino-acid transport, and to determine whether the putative light chain(s) of rBAT, the luminal transporter of the intestine and kidney (disfunctional in cystinuria), also belong(s) to the same permease-related family of lipophilic proteins.

The extracellular domains of glycoproteins rBAT and h4F2hc both include a region of similarity with glycosidases²⁰, but the role of this domain in amino-acid transport is not clear. These glycoproteins may be multifunctional and have other roles beyond enabling surface expression of amino-acid-transport proteins of the AmAT light-chain family. □

Methods

Xenopus oocytes. To remove follicle cells, oocytes were treated for 4 h at room temperature with collagenase (0.83 mg ml⁻¹) in Ca²⁺-free buffer containing (in mM): NaCl, 82.5; KCl, 2; MgCl₂, 1; and HEPES, 10; pH 7.4. Treated oocytes were then kept for at least 24 h at 16 °C in ND96 buffer containing (in mM): NaCl, 96; KCl, 2; MgCl₂, 1; CaCl₂, 1.8; and HEPES, 5; pH 7.4.

Amino-acid uptake and efflux. Oocytes were injected with 33 nl H₂O containing 5 ng of the indicated *in vitro*-transcribed cRNA(s) and incubated in ND96 at 16 °C for 24 h. Amino-acid uptakes were performed in a buffer containing (in mM): NaCl or choline chloride, 100; KCl, 2; MgCl₂, 1; CaCl₂, 1; HEPES, 10; pH 7.4. Tritiated L-amino acids (or ¹⁴C-L-glutamine) were used individually as tracers and the desired final concentrations obtained by addition of unlabelled amino acids. For each test, six oocytes were first pre-incubated for 2 min at 22 °C in amino-acid-free uptake buffer. This was replaced by 100 μ l buffer containing the test amino acid and incubated for 1–15 min. Oocytes were then rapidly washed five times with 3 ml uptake buffer containing 10 mM cold amino acid and distributed to separate vials (total time <1 min). After lysis in 2% SDS, the radioactivity was counted. In preliminary experiments, the time course of the uptake was measured for different amino acids at low and high concentrations to determine the time of linear uptake to be used for subsequent experiments. Results are expressed as mean uptake rates $\pm$ s.e. (pmoles per h per oocyte for groups of six oocytes for Fig. 2a, c and for the mean of 10–12 oocytes from two independent series of experiments for Fig. 3).

For efflux experiments, eight oocytes per test condition were loaded by incubation in uptake buffer containing 1 μ M ³H-L-leucine for 15 min. After washing briefly in buffer containing 1 mM MeAIB, oocytes were transferred to 300 μ l Na⁺-free buffer $\pm$ 1 mM cold L-leucine. Aliquots of 2 μ l were taken at the indicated times.

Labelling of oocytes and immunoprecipitation. Non-injected or cRNA-injected oocytes were incubated in ND96 supplemented with 1 mCi ml⁻¹ [³⁵S]methionine for 48 h at 16 °C. Oocytes were then washed twice with ND96 and lysed in a buffer containing (in mM): Tris/HCl, 50, pH 8.0; NaCl, 120; 0.5% NP40; and protease inhibitors. Extracts were vortexed and centrifuged for 10 min at 12,000 r.p.m. at 4 °C. The supernatant was separated from the pelleted yolk granules and stored at -70 °C. Aliquots of 6 $\times$ 10⁶ c.p.m. were rotated for 16 h at 4 °C with 1 μ g anti-h4F2hc¹⁸ prebound to protein G/protein A agarose. The beads were washed five times with (in mM): Tris/HCl, 20, pH 8.0; NaCl, 100; EDTA, 1; LiCl, 500; 0.5% NP40; and five times with (in mM): Tris/HCl, 20, pH 8.0; NaCl, 100; EDTA, 1; 0.5% NP40. The resulting immunoprecipitates were heated in sample buffer (with or without β -mercaptoethanol) for 5 min at 65 °C.

Immunocytochemistry. Oocytes were fixed with 3% paraformaldehyde in PBS for 4 h at 4 °C, 24 h after injection of the cRNAs. Anti-h4F2hc antibody (6 μ g ml⁻¹) followed by a 1/1,000 dilution of a Cy-3-conjugated goat anti-mouse Ig (Jackson) and a rabbit anti-SPRM1 antibody (P.J.S. and C.B.S., unpublished results) diluted 1/1,000 and followed by a 1/100 dilution of a horseradish peroxidase (HRP)-conjugated donkey anti-rabbit Ig (Jackson) were used on cryosections (thickness, 6 μ m) of the fixed oocytes. The HRP-conjugated secondary antibody was revealed with fluorescein isothiocyanate (FITC)-tyramide conjugate (TSA-Direct, NEN), which was diluted 1/50 in the TSA amplification diluent. Sections were studied by epifluorescence and digitized images were acquired with a VISICAM CCD camera (Visitron) and processed by Image-Pro Plus v3.0 software (Media Cybernetics) (using the same settings for all images with the same antibody).

In vitro translation of cRNA. *In vitro* translation was done using the rabbit reticulocyte lysate system with or without canine pancreatic microsomal membranes. The reaction was done according to the manufacturer's protocol (Promega) using 0.5 μ g cRNA in a final volume of 25 μ l. Half of the microsomal reactions were washed with a twofold volume of 0.2 M sucrose, 10 mM Tris (pH 8.0), and were centrifuged at 125,000g at 4 °C for 5 min. The supernatant was saved and the microsomes were washed a second time using the same amount of wash buffer.

Received 3 June; accepted 16 July 1998.

- Christensen, H. N. Role of amino acid transport and countertransport in nutrition and metabolism. *Physiol. Rev.* **70**, 43–77 (1990).
- Oxender, D. L. & Christensen, H. N. Distinct mediating systems for the transport of neutral amino acids by the Ehrlich cell. *J. Biol. Chem.* **238**, 3686–3699 (1963).
- Devés, R., Chavez, P. & Boyd, C. A. R. Identification of a new transport system (γ^+ L) in human erythrocytes that recognizes lysine and leucine with high affinity. *J. Physiol. (Lond.)* **454**, 491–501 (1992).
- Haynes, B. F. *et al.* Characterization of a monoclonal antibody (4F2) that binds to human monocytes and to a subset of activated lymphocytes. *J. Immunol.* **126**, 1409–1414 (1981).
- MacDonald, H. R., Lees, R. K. & Bron, C. Cell surface glycoproteins involved in the stimulation of interleukin 1-dependent interleukin 2 production by a subline of E14 thymoma cells. I. Functional characterization by monoclonal antibodies. *J. Immunol.* **135**, 3944–3950 (1985).
- Wells, R. G., Lee, W. S., Kanai, Y., Leiden, J. M. & Hediger, M. A. The 4F2 antigen heavy chain induces uptake of neutral and dibasic amino acids in *Xenopus* oocytes. *J. Biol. Chem.* **267**, 15285–15288 (1992).
- Bertran, J. *et al.* Stimulation of system γ^+ -like amino acid transport by the heavy chain of human 4F2 surface antigen in *Xenopus laevis* oocytes. *Proc. Natl Acad. Sci. USA* **89**, 5606–5610 (1992).
- Bröer, S., Bröer, A. & Hamprecht, B. The 4F2hc surface antigen is necessary for expression of system L-like neutral amino acid-transport activity in C6-BU-1 rat glioma cells: evidence from expression studies in *Xenopus laevis* oocytes. *Biochem. J.* **312**, 863–870 (1995).
- Bröer, S., Bröer, A. & Hamprecht, B. Expression of the surface antigen 4F2hc affects system-L-like neutral-amino-acid-transport activity in mammalian cells. *Biochem. J.* **324**, 535–541 (1997).
- Hemler, M. E. & Strominger, J. L. Characterization of antigen recognized by the monoclonal antibody (4F2): different molecular forms on human T and B lymphoblastoid cell lines. *J. Immunol.* **129**, 623–628 (1982).
- Lüscher, B., Rousseaux, M., Lees, R., MacDonald, H. R. & Bron, C. Cell surface glycoproteins involved in the stimulation of interleukin 1-dependent interleukin 2 production by a subline of E14 thymoma cells. II. Structure, biosynthesis, and maturation. *J. Immunol.* **135**, 3951–3957 (1985).
- Gaugitsch, H. W., Prieschl, E. E., Kalthoff, F., Huber, N. E. & Baumrucker, T. A novel transiently expressed, integral membrane protein linked to cell activation. *J. Biol. Chem.* **267**, 11267–11273 (1992).
- Spindler, B., Mastroberardino, L., Custer, M. & Verrey, F. Characterization of early aldosterone-induced RNAs identified in A6 kidney epithelia. *Pflügers Arch.* **434**, 323–331 (1997).
- Isnard, A. D., Thomas, D. & Surdin-Kerjan, Y. The study of methionine uptake in *Saccharomyces cerevisiae* reveals a new family of amino acid permeases. *J. Mol. Biol.* **262**, 473–484 (1996).
- Hofmann, K. & Stoffel, W. Profilegraph: an interactive graphical tool for protein sequence analysis. *Comp. Appl. Biosci.* **8**, 331–337 (1992).
- Palacin, M. A new family of proteins (rBAT and 4F2hc) involved in cationic and zwitterionic amino acid transport: a tale of two proteins in search of a transport function. *J. Exp. Biol.* **196**, 123–137 (1994).
- Teixeira, S., Di Grandi, S. & Kuhn, L. C. Primary structure of the human 4F2 antigen heavy chain predicts a transmembrane protein with a cytoplasmic NH2 terminus. *J. Biol. Chem.* **262**, 9574–9580 (1987).
- Eisenbarth, G. S., Haynes, B. F., Schroer, J. A. & Fauci, A. S. Production of monoclonal antibodies reacting with peripheral blood mononuclear cell surface differentiation antigens. *J. Immunol.* **124**, 1237–1244 (1980).
- Geering, K., Theulaz, L., Verrey, F., Häuptle, M. T. & Rossier, B. C. A role for the β -subunit in the expression of functional Na^+/K^+ -ATPase in *Xenopus* oocytes. *Am. J. Physiol.* **257**, C851–C858 (1989).
- Wells, R. G. & Hediger, M. A. Cloning of a rat kidney cDNA that stimulates dibasic and neutral amino acid transport and has sequence similarity to glucosidases. *Proc. Natl Acad. Sci. USA* **89**, 5596–5600 (1992).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank L. Kühn and M. Palacin for providing antibody and the h4F2hc cDNA. H. Murer and J. Biber for the *Xenopus* oocytes, and all four for helpful discussions. We also thank C. Gasser for artwork and L. Dionne, H. J. Kim and L. Kläusli for technical assistance. F.V.'s laboratory is supported by the Swiss NSF and the work in C.B.S.'s laboratory is supported by an NIH grant.

Correspondence and requests for materials should be addressed to F.V. (e-mail: verrey@physiol.unizh.ch). Amino-acid sequences have been deposited in GenBank/EMBL under accession numbers: ASUR4, Y12716; E16, AF077866; SPRM-1, L25068.

IKAP is a scaffold protein of the I κ B kinase complex

Lucie Cohen*, William J. Henzel† & Patrick A. Baeuerle*

* Tularik Inc., Two Corporate Drive, South San Francisco, California 94080, USA
† Genentech Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

The transcription factor NF- κ B coordinates the activation of numerous genes in response to pathogens and pro-inflammatory cytokines, and is, therefore, vital in the development of acute and chronic inflammatory diseases^{1–6}. NF- κ B is activated by phosphorylation of its inhibitory subunit, I κ B- α (ref. 7), on serine residues 32 and 36 by cytokine-activated I κ B kinases (IKKs); this phosphorylation precedes rapid degradation of I κ B^{8–11}. IKK- α and IKK- β isozymes are found in large complexes of relative molecular mass 700,000–900,000 (M_r 70K–90K), but little is known about other components that organize and regulate these complexes^{12–17}. IKK- α was independently discovered as a NF- κ B-inducing kinase¹⁸ (NIK)-associated protein in a yeast two-hybrid screen¹⁹, and IKK- β was also identified by homology screening²⁰. It is, however, unknown whether NIK is part of the IKK complex. Here we isolate large, interleukin-1-inducible IKK complexes that contain NIK, IKK- α , IKK- β , I κ B- α , NF- κ B/RelA and a protein of M_r 150K. This latter component is a new protein, termed IKK-complex-associated protein (IKAP), which can bind NIK and IKKs and assemble them into an active kinase complex.

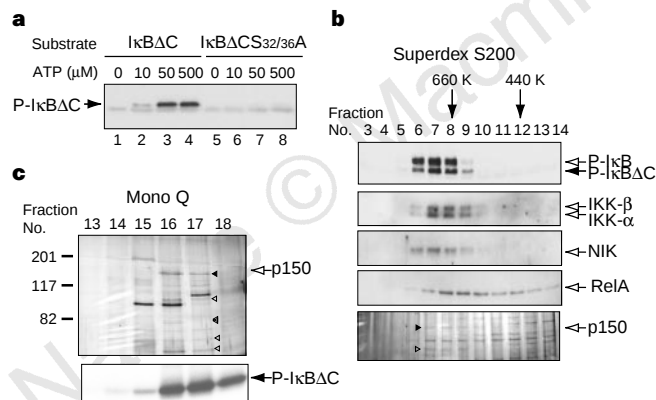


Figure 1 Purification and composition of large IKK complexes. **a**, An I κ B-affinity-purified fraction (0.5 μ l) was tested in kinase reactions against two forms of I κ B in the presence of the indicated concentrations of cold ATP. The position of phosphorylated I κ B (P-I κ B Δ C) is indicated on the immunoblot section (the crossreacting band seen in all lanes corresponds to unphosphorylated I κ B- α). **b**, I κ B-affinity-purified proteins were fractionated by Superdex S200 and were tested for S32/36-specific IKK activity (top), IKK- α and IKK- β protein (second panel), NIK (third panel) and RelA (fourth panel) using specific polyclonal rabbit antisera (provided by M. Roth and J. Woronicz), or were analysed by silver staining (bottom). The peak distribution of 660K and 440K size exclusion markers are shown. The positions of the different polypeptides are indicated by arrows on the right, including the positions of phosphorylated endogenous I κ B- α (top panel, open arrow) and recombinant I κ B Δ C (filled arrow). Within the bottom panel, the filled and open arrowheads point to silverstained polypeptides with apparent relative molecular masses of 150K and 105K respectively. **c**, MonoQ fractions were analysed by silver staining for p150 (top) and by radioactive kinase assay (bottom). Polypeptides of 100K, 82K, 80K, 65K and 58K co-eluting with IKK activity are also shown (open arrowheads). Molecular mass standards are on the left. Filled and open arrowheads indicate the positions of additional polypeptides co-purifying with IKK activity.

We show that IKAP is a scaffold protein and a regulator for three different kinases involved in pro-inflammatory cytokine signalling.

Native IKK complexes were purified >10,000-fold by a rapid I κ B-substrate-affinity chromatography protocol from interleukin-1 (IL-1)-stimulated 293 cells that stably overexpress the IL-1 receptor type I (see Methods). We detected IKK activity in I κ B-affinity eluates by western blot, using a polyclonal antiserum that specifically recognizes the S32/S36-phosphorylated form of I κ B- α (Fig. 1a). We detected I κ B phosphorylation when a minimal ATP concentration of 10 μ M was used (Fig. 1a, lanes 2–4), and not with a S32/36A-I κ B mutant (Fig. 1a, lanes 6–8). I κ B-affinity-purified proteins were fractionated by gel filtration and analysed by western blot using several specific rabbit polyclonal antisera (Fig. 1b). As detected by the I κ B-phosphoserine antiserum, IKK activity eluted in fractions of 700K–900K (Fig. 1c, top), corresponding to the previously reported size of native IKK complexes^{16,17}. A more slowly migrating form of phosphorylated I κ B (Fig. 1c) was also detected which migrated with endogenous I κ B- α (data not shown), indicating that IKK complexes may have at least two binding sites for I κ B. A mixture of IKK α - and IKK β -specific polyclonal antibodies detected a doublet band of ~80K that eluted together with the peak of IKK activity (Fig. 1c). This identical elution profile indicates that native IKK complexes may contain both IKK- α and IKK- β as heterodimers^{16,17,20}. NIK was also detected and co-eluted nearly with the two IKKs, and even better with IKK activity (Fig. 1c). In contrast, RelA was present in the IKK peak but showed a more heterogeneous elution profile. These data indicate that NIK and, presumably, RelA-I κ B α are also components of large IKK complexes.

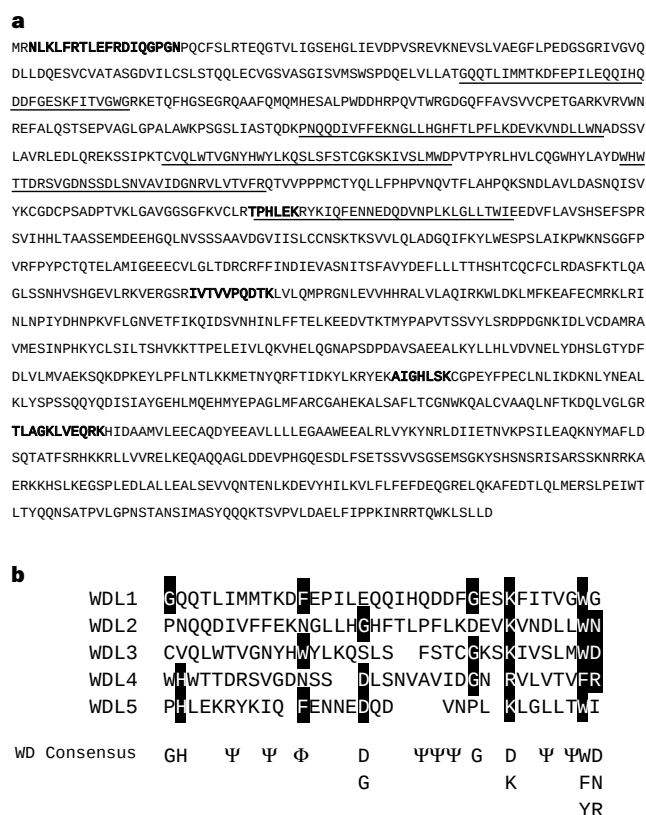


Figure 2 Characteristics of p150/IKAP. **a**, Primary structure of p150 as deduced from the full-length cDNA. Five WD-like repeats are underlined. Microsequenced peptides are in bold. **b**, WD-like repeats were aligned with the consensus sequence of WD repeats²². Conserved positions are highlighted in black. Ψ , hydrophobic amino acid; Φ , aromatic amino acid.

showing that endogenous NIK and IKK- β exist in a complex with endogenous IKAP (Fig. 3b). We precipitated endogenous IKAP from untransfected 293/IL-1-receptor cells using antibodies that recognize an amino-terminal peptide sequence (Fig. 3b, top; compare lanes 1, 2). When tested with a polyclonal anti-NIK antiserum, IKAP immunoprecipitate contained a 100K protein corresponding to NIK (Fig. 3b, centre panel, lane 2); this protein was undetected by the control serum (lane 1). Equal amounts of NIK were found in a complex with IKAP before and after cytokine stimulation (data not shown). Endogenous IKK- β was also present in a complex with endogenous IKAP (Fig. 3b, lower panel, lane 3). This interaction was cytokine-dependent. After 10 min of stimulation with IL-1,

most of the IKK- β was no longer bound to IKAP (lane 4). Immunoprecipitation with IKK- β polyclonal antiserum detected an IKK- β signal of similar intensity to that obtained with the anti-IKAP antibody (lanes 5, 6), indicating that most of the cellular IKK- β may be associated with IKAP. The equivalent experiment could not be performed with IKK- α because of the presence of a background band obscuring detection of the kinase.

Interactions between IKAP and the kinases (IKK- α , IKK- β and NIK) did not require other proteins of vertebrate cells. This was evident from a pull-down experiment using recombinant proteins that were overexpressed separately and affinity-purified from insect cells. All three kinases were co-immunoprecipitated in a complex with IKAP and were undetected in control reactions lacking IKAP (Fig. 3c). The relative amounts of kinases precipitated by IKAP were similar to those observed in *in vivo* overexpression experiments (Fig. 3a). Detection of IKK- β over a background band was confirmed using an anti-IKK β polyclonal antibody (Fig. 3c, lane 8, lower panel). The high level of baculovirus-mediated expression and the affinity purification of the recombinant proteins make it unlikely that homologous insect proteins mediated these interactions. The data indicate that IKAP can directly and independently bind NIK, IKK- α and IKK- β .

With these properties, IKAP may act as a scaffold that assembles IKKs with their upstream regulatory kinase NIK. In support of this idea, we found that IKAP can bind NIK and IKKs through separate domains (Fig. 3d). We tested a series of amino- and carboxy-terminal truncation mutants of IKAP for their ability to bind NIK, IKK- α and IKK- β after being overexpressed in 293 cells. All IKAP truncations showed the expected molecular size in SDS gels and comparable expression levels (data not shown). IKAP residues 441–600, containing the fifth WD-like repeat, were required for interaction with NIK (Fig. 3d). In contrast, interaction with IKK- α and IKK- β required two C-terminal overlapping sequences. Binding to IKK- α depended on residues 681–920, whereas binding to IKK- β required residues 801–920 only of IKAP. This difference in binding-domain sizes may explain why IKK- α shows a higher affinity for IKAP than does IKK- β .

If IKAP is a functional component of the IKK complex, its overexpression should inhibit cytokine signalling by titrating out each individual component of the IKK complex. IKAP expression indeed inhibited, in a dose-dependent manner, NF- κ B-dependent luciferase gene expression induced by tumour-necrosis factor- α (Fig. 4a). IKAP expression also inhibited IL-1-induced NF- κ B-dependent luciferase gene expression, down to levels seen in the absence of cytokine stimuli (Fig. 4b). Transdominant-negative NIK²⁰ and TRAF6 Δ (ref. 23) seemed to be more potent inhibitors of gene expression than IKAP, but immunoblotting using the same epitope-specific antibody showed that the transdominant mutants were expressed to a much higher level than IKAP (inserts in Fig. 4a; data not shown). Overexpression of IKAP in 293 cells alone caused a four- to fivefold, reproducible activation of a κ B-dependent luciferase reporter gene controlled by the ELAM promoter (Fig. 4a) at low concentrations of expression vector. IKAP effects were specific to the NF- κ B pathway and did not affect the interferon- γ -induced activation of a luciferase reporter controlled by cyclic-AMP-response element (CRE) and STAT transcription factor motifs (Fig. 4c).

To determine whether IKAP can assemble active IKK–NIK complexes, we overexpressed IKAP with various combinations of NIK, IKKs, and their respective kinase-inactive mutants (Fig. 5b). IKAP immunoprecipitates were analysed for the relative abundance of kinases (Fig. 5a, top) and for kinase activity of IKK (Fig. 5a, centre). As shown above (Fig. 3a), IKAP precipitated distinct amounts of the individual kinases (Fig. 5a, lanes 2–4). The IKK- β band was visible only upon longer exposure of the blot. Intriguingly, none of the complexes containing IKAP and a single kinase showed a significant level of IKK activity. In the complex with IKAP, only NIK appeared to be phosphorylated (Fig. 5a, lane 4). When

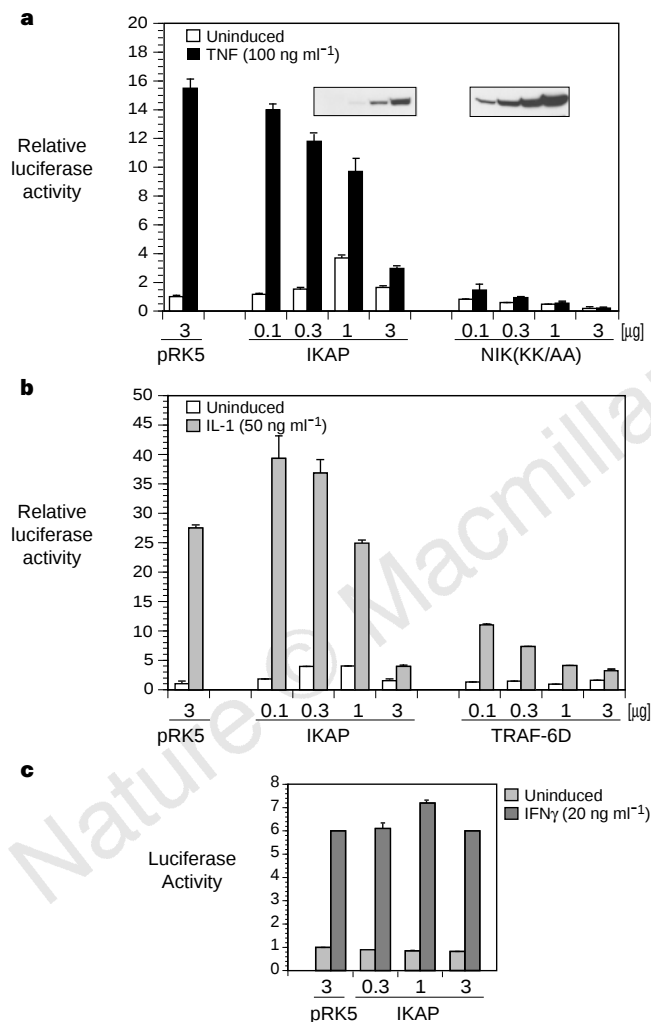
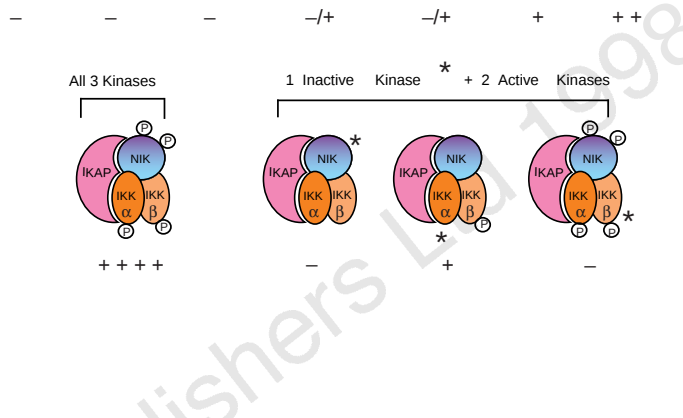


Figure 4 IKAP expression affects IL-1 and tumour-necrosis factor (TNF) signalling pathways. **a**, 293 cells were transiently transfected with 0.5 μg of an ELAM reporter construct and 1 μg of β -galactosidase expression plasmid¹⁹, together with variable amounts of vector control (pRK5) and the indicated amounts of Flag-tagged IKAP or NIK(KK429,430AA) expression vectors. 36 h after transfection, cells were either left untreated (open bars), or treated for 6 h with TNF (filled bars) at 100 ng ml⁻¹ before being collected. Luciferase activity was normalized to levels of β -galactosidase activity. The data shown are the mean values from one representative experiment (of at least three) in which each transfection was performed in duplicate. Standard deviations from the mean are shown. Inserts show the level of overexpressed IKAP (left) and kinase-inactive NIK (right) detected by western blotting using anti-Flag IgG. **b**, As for **a**, except that 293/IL-1-receptor cells were transfected with IKAP or TRAF6 Δ expression vectors. **c**, As in **a**, except that a luciferase reporter construct under the control of STAT- and CRE-binding sites was transfected and 20 ng ml⁻¹ interferon- γ was used for induction of 293 cells 30 h after transfection.



band also present in extract lacking IKAP (lane 12). Lower asterisk, phosphoprotein band which co-migrated with a degradation product of IKAP. The amount of Coomassie blue-stained IKAP in each immunoprecipitate is shown (bottom). **b**, The different complexes assembled by IKAP *in vivo* are shown. The IKK activity of the complex containing all three wild-type kinases is set to 100% (++++). The dashed lines indicate the possibility that IKKs bind as homodimers to IKAP. Phosphorylation of kinases is shown by a circled P. Asterisks indicate point-mutated inactive forms of the kinases.

Our results are, to our knowledge, the first to show that NIK and IKAP are components of cytokine-inducible IKK-containing complexes. The presence of NIK in IKK complexes is consistent with NIK's affinity for IKK- α , seen in yeast two-hybrid and 293 cell overexpression systems^{19,20}, and with the observation that over-expressed NIK stimulates the activity of both IKK- α and IKK- β (refs 24, 25). We also show that NIK is an essential regulatory component of IKK complexes in our reconstitution experiments, in which wild-type NIK enhanced the kinase activity of IKKs whereas a kinase-inactive NIK mutant completely inhibited the activity of IKKs in the IKAP complex.

The interaction of IKAP with three kinases is reminiscent of the function of Ste5, a yeast protein which acts as a scaffold protein by binding and assembling kinases of the MAP kinase cascade into a multikinase complex^{26–28}. So far, there is no evidence for the existence of mammalian Ste5 homologues or analogues. We found that IKAP can interact concurrently with three kinases through distinct domains of IKAP. Although the binding of IKK- α and IKK- β to IKAP may be mutually exclusive, because they bind overlapping domains, an IKK heterodimer may require only one binding site within IKAP. Our results indicate that IKAP not only assembles kinases to form a large IKK complex, but is also required to impose tight control by NIK over the IKKs and to allow for a strong mutual regulation of the three kinases.

NATURE | VOL 395 | 17 SEPTEMBER 1998

only the fifth WD-like repeat was sufficient for binding of NIK. This would leave the first four WD-like repeats, which may act as protein-protein interaction motifs²², free for additional interactions of IKAP with, for example, upstream regulators of NIK. □

Methods

Purification of IKK complexes. For preparation of the IκB-affinity matrix, recombinant IκB proteins were purified from cultures of BL21 cells transformed with Flag-tagged IκBΔC expression vector¹⁹ and lysed in bacterial lysis buffer (50 mM HEPES, pH 7.9, 300 mM NaCl, 0.1% Nonidet P-40, 2 mM dithiothreitol (DTT), 1 mM EDTA, 1 mM sodium metabisulphite) containing 0.4 mM phenylmethylsulphonylfluoride (PMSF) and 3.5 mg ml⁻¹ lysozyme. Extracts were cleared and incubated with the anti-Flag monoclonal antibody M2 coupled to Sepharose beads (Eastman Kodak) for 1 h at 4 °C. After intensive high-salt washes, the binding capacity of the affinity matrix was ~1 mg IκBΔC per ml of beads as estimated by SDS-PAGE. Suspension cultures of 293/IL-1-receptor cells²⁹ (100 l, 1.2 × 10⁶ cells per ml) were induced with 200 ng ml⁻¹ recombinant human IL-1 (Genentech) for 5 min at 37 °C. Total extracts were prepared by lysing cells in ice-cold buffer A (50 mM HEPES, pH 7.6, 250 mM NaCl, 0.1% (w/v) Nonidet P-40, 2 mM MgCl₂, 1 mM EDTA, 10% (v/v) glycerol, 0.5 mM DTT) containing 1 mM NaF, 1 mM Na₃VO₄, 20 mM β-glycerophosphate, 5 mM *p*-nitrophenylphosphate, 1 mM benzamide, 1 mM sodium metabisulphate and a protease inhibitor cocktail (Complete, Boehringer) for 30 min on ice. Cell extracts were cleared by ultracentrifugation at 100,000 g for 2–5 h and the supernatants (S-100 fractions) were pooled. S-100 fractions were incubated with IκBΔC-affinity matrix (~1 ml beads per g total protein) for 2 h at 4 °C with gentle rocking. Beads were extensively washed (5 × 20 ml buffer A per ml bed volume), and collected in an Econo-column (Bio-Rad). Bound proteins were eluted with five volumes of buffer A containing 100 mM NaCl and 0.3 mg ml⁻¹ Flag epitope peptide (Eastman Kodak). Eluates were loaded on a MonoQ column (1 ml) and proteins were eluted using a linear 0.1–1 M NaCl gradient in buffer A. Samples of Flag eluates or active MonoQ fractions (450–550 nM NaCl fractions) were analysed by gel filtration on Superdex S200 (2.5 ml) equilibrated with buffer A containing 150 mM NaCl using the Smart System (Pharmacia). Fractions were analysed by silver staining, western blotting and kinase assays for protein composition and monitoring IKK activity.

Sequencing and cDNA cloning of IKAP. Polypeptides in active MonoQ fractions were purified by preparative SDS-PAGE and transferred onto a PVDF membrane (Applied Biosystems). Peptides from the excised 150K band were generated by *in situ* tryptic proteolysis, separated on a capillary gradient high-performance liquid chromatography (HPLC) system³⁰, and analysed by mass spectrometry and automated protein sequencing on a model 494 Applied Biosystems sequencer. Degenerate oligonucleotide primers were designed from microsequenced peptides and used to amplify a 600-bp fragment by the polymerase chain reaction (PCR). Human HeLa cell libraries were screened with the 600-bp-fragment probe and two overlapping clones of 3.8 kb and 1.6 kb were isolated. The full-length cDNA clone was reconstituted by *NotI*–*XbaI* restriction digest of the inserts and religation, and then subcloned into the pRK5 vector in-frame with an N-terminal Flag-epitope-tag sequence or a C-terminal Myc-epitope-tag sequence. Truncation mutants of IKAP were generated using internal restriction sites and/or PCR. All mutations were confirmed by DNA sequencing.

Kinase assays and immunoprecipitation. Kinase reactions were done in solution with 0.5–5 μl Flag eluates or fractions from Superdex S200 and MonoQ columns and 0.5–2 μg IκBΔC substrate in 20–30 μl kinase buffer (20 mM HEPES, pH 7.6, 20 mM MgCl₂, 2 mM DTT, 0.1 mM Na₃VO₄, 20 mM β-glycerophosphate and 20 mM *p*-nitrophenylphosphate) containing 10–500 μM cold ATP or 20 μM ATP containing 5 μCi [γ-³²P]ATP per reaction (1,000 Ci mmol⁻¹, NEN-Dupont). Non-radioactive *in vitro* kinase reactions were analysed by immunoblotting using a rabbit polyclonal antibody recognizing a synthetic IκB-α peptide phosphorylated on serines 32 and 36 (provided by C. Lehel). A donkey anti-rabbit antibody coupled to horseradish peroxidase was used as secondary antibody, and the complex was visualized using the enhanced chemiluminescence (ECL) western blotting detection system (Amersham). Immunoprecipitations of proteins from transfected and non-transfected 293 cells were performed as described^{19,20}. Antisera against an

IKAP N-terminal peptide (NLKLFRTLEFRDIQGPGE) were produced by the PolyQuick Antibodies Program (Zymed Lab). Polyclonal anti-Flag and anti-Myc antibodies (IgG) were from Santa Cruz Biotechnology, and monoclonal anti-Flag antibody (mIgG) was from Eastman Kodak.

Expression vectors and reporter gene assays. Mammalian expression vectors encoding NIK, IKK-α, IKK-β, p65/RelA and p50 were provided by C. Régnier, J. Woronicz and V. Baichwal. 293 and 293/IL-1-receptor cells were transfected by the calcium phosphate precipitation method using the Profection Kit (Promega). Reporter gene assays were performed as described^{19,20}. The STAT/CRE-responsive luciferase reporter construct was provided by T. Hoey.

Purification of IKAP from insect cells. IKAP cDNA was subcloned into pBlueBacHis2C vector (Invitrogen) in-frame with N-terminal polyhistidine and Xpress epitope-tag sequences. Recombinant IKAP protein was purified from Sf21 cells (~5 × 10⁶) using 400 μl Ni-nitrilotriacetic acid agarose bead suspension (Qiagen) according to the manufacturer's instructions. Purified Flag-tagged NIK, IKK-α and IKK-β from insect cells were provided by M. Ayres and S. Li. Protein concentrations were estimated by Coomassie blue staining.

Received 29 May; accepted 4 August 1998.

1. Siebenlist, U., Brown, K. & Franzoso, G. In *Inducible Gene Expression* Vol. 1 (ed. Baueerle, P. A.) 93–141 (Birkhaeuser, Boston, 1995).
2. Verma, I. M., Stevenson, J. K., Schwarz, E. M., Van Antwerp, D. & Miyamoto, S. Rel/NF-κB/IκB family: intimate tales of association and dissociation. *Genes Dev.* **9**, 2723–2735 (1995).
3. Baueerle, P. A. & Baltimore, D. NF-κB: ten years after. *Cell* **87**, 13–20 (1996).
4. Baldwin, A. S. The NF-κB and I-κB proteins: new discoveries and insights. *Annu. Rev. Immunol.* **14**, 649–681 (1996).
5. Thanos, A. M. & Maniatis, T. NF-κB: a lesson in family values. *Cell* **80**, 529–532 (1995).
6. Baueerle, P. A. & Baichwal, V. R. NF-κB as frequent target for immunosuppressive and anti-inflammatory molecules. *Adv. Immunol.* **6**, 111–137 (1997).
7. Baueerle, P. A. & Baltimore, D. IκB: a specific inhibitor of the NF-κB transcription factor. *Science* **242**, 540–546 (1988).
8. Brockman, J. A. *et al.* Coupling of a signal response domain in IκB to multiple pathways for NF-κB activation. *Mol. Cell. Biol.* **15**, 2809–2818 (1995).
9. Brown, K., Gerstberger, S., Carlson, L., Franzoso, G. & Siebenlist, U. Control of IκBα proteolysis by site-specific, signal-induced phosphorylation. *Science* **267**, 1485–1491 (1995).
10. Traencker, E. B.-M. *et al.* Phosphorylation of human IκB-α on serine 32 and 36 controls IκB-α proteolysis and NF-κB activation in response to diverse stimuli. *EMBO J.* **14**, 2876–2883 (1995).
11. DiDonato, J. A. *et al.* Mapping of the inducible IκB phosphorylation sites that signal its ubiquitination and degradation. *Mol. Cell. Biol.* **16**, 1295–1304 (1996).
12. Maniatis, T. Catalysis by a multiprotein IκB kinase complex. *Science* **278**, 818–819 (1997).
13. Stancovski, I. & Baltimore, D. NF-κB activation: the IκB kinase revealed? *Cell* **91**, 299–302 (1997).
14. Baueerle, P. A. Pro-inflammatory signaling: last pieces in the NF-κB puzzle? *Curr. Biol.* **8**, 19–22 (1998).
15. DiDonato, J. A., Hayakawa, M., Rothwarf, D. M., Zandi, E. & Karin, M. A cytokine-responsive IκB kinase that activates the transcription factor NF-κB. *Nature* **388**, 548–554 (1997).
16. Zandi, E., Rothwarf, D. M., Delhase, M., Hayakawa, M. & Karin, M. The IκB kinase complex (IKK) contains two kinase subunits, IKKα and IKKβ, necessary for IκB phosphorylation and NF-κB activation. *Cell* **91**, 243–252 (1997).
17. Mercurio, F. *et al.* IKK-1 and IKK-2: cytokine-activated IκB kinases essential for NF-κB activation. *Science* **278**, 860–866 (1997).
18. Malinin, N. L., Boldin, M. P., Kovalenko, A. V. & Wallach, D. MAP3K-related kinase involved in NF-κB induction by TNF, CD95 and IL-1. *Nature* **385**, 540–544 (1997).
19. Régnier, C. H. *et al.* Identification and characterization of an IκB kinase. *Cell* **90**, 373–383 (1997).
20. Woronicz, J. D., Gao, X., Cao, Z., Rothe, M. & Goeddel, D. V. IκB kinase-β: NF-κB activation and complex formation with IκB kinase-α and NIK. *Science* **278**, 866–869 (1997).
21. Yajima, H., Tokunaga, M., Nakayama-Murayama, A. & Hishinuma, F. Characterization of IKI1 and IKI3 genes conferring pGKL killer sensitivity on *Saccharomyces cerevisiae*. *Biosci. Biotech. Biochem.* **61**, 704–709 (1997).
22. Neer, E. J., Schmidt, C. J., Nambudripad, R. & Smith, T. F. The ancient regulatory-protein family of WD-repeat proteins. *Nature* **371**, 297–300 (1994).
23. Cao, Z., Xiong, J., Takeuchi, M., Kurama, T. & Goeddel, D. V. TRAF-6 is a signal transducer for interleukin-1. *Nature* **383**, 443–446 (1996).
24. Lei, L., Cao, Z. & Goeddel, D. V. NF-κB-inducing kinase activates IKK-α phosphorylation of Ser-176. *Proc. Natl Acad. Sci. USA* **95**, 3792–3797 (1998).
25. Nakano, H. *et al.* Differential regulation of the IκB kinase α and β by two upstream kinases, NF-κB-inducing kinase and mitogen-activated protein kinase/ERK kinase kinase-1. *Proc. Natl Acad. Sci. USA* **95**, 3527–3542 (1998).
26. Choi, K. Y., Scatterberg, B., Lyons, D. M. & Ellison, E. A. Ste5 tethers multiple protein kinases in the MAP kinases in the MAP kinase cascade required for mating of *S. cerevisiae*. *Cell* **78**, 499–512 (1994).
27. Printen, J. A. & Sprague, G. F. Jr Protein-protein interactions in the yeast pheromone response pathway: Ste5p interacts with all members of the MAP kinase cascade. *Genetics* **138**, 609–619 (1994).
28. Marcus, S., Polverino, A., Barr, M. & Wigler, M. Complexes between Ste5 and components of the pheromone-responsive mitogen-activated protein kinase module. *Proc. Natl Acad. Sci. USA* **91**, 7762–7766 (1994).
29. Cao, Z., Henzel, W. J. & Gao, X. IRAK: a kinase associated with the interleukin-1 receptor. *Science* **271**, 1128–1131 (1996).
30. Henzel, W. J. & Stults, J. T. Reversed phase isolation of peptides. *Curr. Prot. Protein Sci.* **1** 11.6.1–11.6.14 (1995).

Acknowledgements. We thank J. Woronicz, C. Régnier, M. Ayres, S. Li, C. Lehel and T. Hoey for providing reagents; K. Williamson for DNA sequencing; and C. Béraud, M. Rothe, Z. Cao and D. Goeddel for discussions throughout the project and comments on the manuscript.

Correspondence and requests for materials should be addressed to P.A.B. (e-mail: baeuerle@tularik.com). The IKAP cDNA sequence has been deposited with GenBank (accession number AF044195).

IKK- γ is an essential regulatory subunit of the I κ B kinase complex

David M. Rothwarf^{*†}, Ebrahim Zandi^{*†}, Gioacchino Natoli^{*} & Michael Karin^{*}

^{*} Laboratory of Gene Regulation and Signal Transduction, Department of Pharmacology, University of California San Diego, 9500 Gilman Drive, La Jolla, California 92093-0636, USA

[†] These authors contributed equally to this work

Pro-inflammatory cytokines activate the transcription factor NF- κ B by stimulating the activity of a protein kinase that phosphorylates I κ B, an inhibitor of NF- κ B¹⁻⁵, at sites that trigger its ubiquitination and degradation. This results in the nuclear translocation of freed NF- κ B dimers and the activation of transcription of target genes^{6,7}. Many of these target genes code for immunoregulatory proteins^{8,9}. A large, cytokine-responsive I κ B kinase (IKK) complex has been purified and the genes encoding two of its subunits have been cloned^{1,2,5}. These subunits, IKK- α and IKK- β , are protein kinases whose function is needed for NF- κ B activation by pro-inflammatory stimuli. Here, by using a monoclonal antibody against IKK- α , we purify the IKK complex to homogeneity from human cell lines. We find that IKK is composed of similar amounts of IKK- α , IKK- β and two other polypeptides, for which we obtained partial sequences. These polypeptides are differentially processed forms of a third subunit, IKK- γ . Molecular cloning and sequencing indicate that IKK- γ is composed of several potential coiled-coil motifs. IKK- γ interacts

preferentially with IKK- β and is required for the activation of the IKK complex. An IKK- γ carboxy-terminal truncation mutant that still binds IKK- β blocks the activation of IKK and NF- κ B.

We previously purified the IKK complex from HeLa cells treated with tumour-necrosis factor (TNF)¹. The complex contains two catalytic subunits, IKK- α and IKK- β , of relative molecular mass 85,000 and 87,000 (M_r 85K and 87K), respectively. These subunits have a kinase domain in their amino-terminal portion and protein-interaction motifs, including a leucine zipper and a helix-loop-helix (HLH)¹⁻⁵, in their C-terminal region¹⁻⁵. IKK- α and IKK- β are rapidly activated by TNF and interleukin-1 (IL-1) and are necessary for NF- κ B activation^{1,2,5}. IKK activity depends on its phosphorylation, as it is inactivated by protein phosphatase 2A (ref. 1). The exact subunit whose dephosphorylation causes loss of IKK activity is not yet known. IKK- α/β can be phosphorylated and activated by overexpressed NF- κ B-inducing kinase (NIK)¹⁰ or by MEK kinase-1 (MEKK-1)¹¹, but the physiological role of NIK and MEKK-1 in IKK activation by pro-inflammatory cytokines is not clear¹². The composition of the IKK complex and the function of its various subunits needed to be determined.

To purify IKK to homogeneity, we prepared monoclonal antibodies specific for IKK- α ; one of these antibodies is very efficient in precipitating IKK activity. Using it as an immunoaffinity reagent, we found that, in both HeLa and Jurkat cells, the IKK complex contained nearly equal amounts of IKK- α and IKK- β (Fig. 1). Furthermore, passing partially purified IKK fractions through the anti-IKK α column resulted in quantitative recovery of IKK- β (Fig. 1a), with which this antibody does not crossreact (our unpublished results). Large-scale purification of IKK, by combining the previous method¹ with immunoaffinity chromatography on immobilized anti-IKK α , showed that, in addition to IKK- α/β , purified IKK contained two polypeptides of M_r 50K and 52K (Fig. 1b). Similar results were obtained by immunoprecipitation

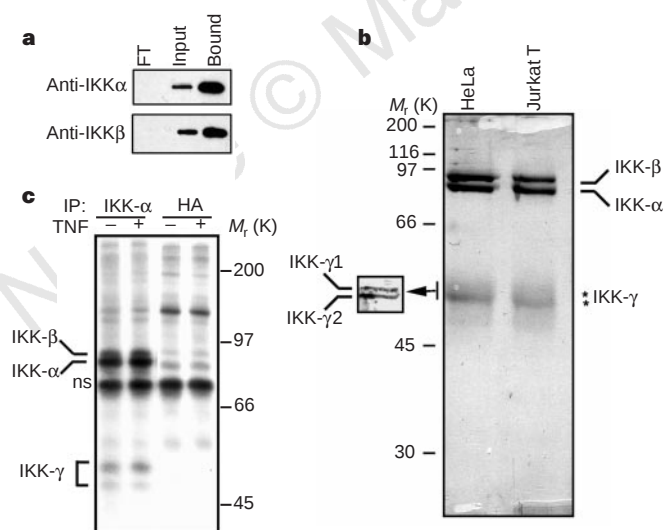


Figure 1 Purification of the IKK complex and identification of the IKK- γ subunits. **a**, Partially purified HeLa cell extracts were passed through an anti-IKK α immunoaffinity column. The input, flowthrough (FT) and bound fractions were separated by SDS-PAGE and their content of IKK- α and IKK- β was assessed by immunoblotting. **b**, The purified IKK complex from HeLa or Jurkat cells was separated by SDS-PAGE and stained with colloidal blue. The positions of the different subunits are indicated. The inset shows a portion of a gel run for a longer time in which better separation of IKK- γ 1 and IKK- γ 2 (asterisked to right) is seen. **c**, 293 cells were labelled with ³⁵S for 5 h, followed by incubation with or without TNF. Lysates were immunoprecipitated (IP) with either anti-IKK α or anti-HA (used as a control) antibodies. After extensive washing the immune complexes were separated by SDS-PAGE and visualized by autoradiography. ns, nonspecific band.

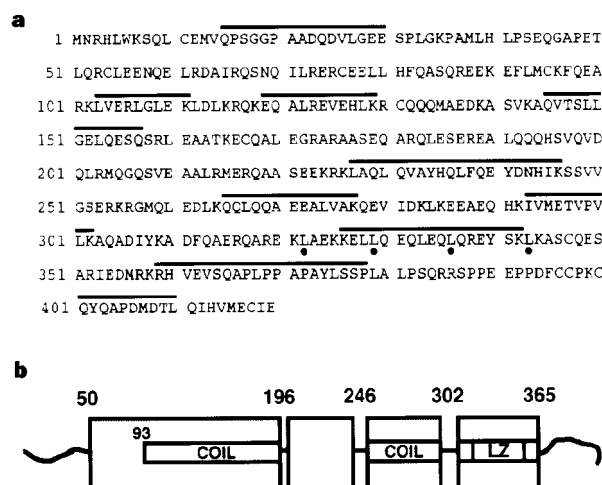


Figure 2 Primary and secondary structure of IKK- γ . **a**, The amino-acid sequence of the complete IKK- γ ORF. Peptide sequences obtained by microsequencing are overlined and the leucines of the leucine zipper are indicated by black dots. **b**, Secondary-structure prediction for IKK- γ . The boxes indicate α -helical regions. Coil, coiled-coil regions; LZ, leucine-zipper motif (which is a coiled-coil). The amino-acid positions that mark the approximate boundaries of these motifs are indicated.

of ^{35}S -labelled cell lysates (Fig. 1c). TNF treatment did not change the relative amounts of these polypeptides. Microsequencing of these polypeptides indicated that they are either differentially processed or differentially initiated forms of the same protein (see Methods), which we named IKK- γ . Screening of Genebank for sequences that are identical or similar to these peptides revealed two overlapping expressed sequence tags (ESTs). A complementary DNA probe corresponding to these ESTs was used to isolate a cDNA clone containing the entire IKK- γ open reading frame (ORF) (Fig. 2a). The predicted IKK- γ polypeptide is 419 amino acids long. Sequence analysis shows that IKK- γ is a new protein and is glutamine-rich. Programs for secondary-structure prediction^{13–15} indicate that the prominent features of IKK- γ are two extended coiled-coil motifs and a leucine zipper (Fig. 2b). Such motifs are likely to be involved in protein–protein interactions.

To confirm that the cloned IKK- γ protein interacts with IKK- α / β subunits in cells, we transfected an expression vector for N-terminally haemagglutinin (HA)-tagged IKK- γ into HeLa cells, together with expression vectors for either Flag-tagged IKK- α or Flag-tagged IKK- β . Transfected cell lysates were immunoprecipitated with anti-HA antibody and then immunoblotted with anti-Flag antibody. This analysis confirmed that IKK- γ interacts efficiently with either IKK- α or IKK- β (Fig. 3a). Similar results were obtained when the immunoprecipitating antibody was directed to the Flag epitope and the immune complexes were immunoblotted with anti-HA antibody, which recognizes the epitope on IKK- γ . Immunoprecipitation of transiently expressed HA–IKK γ resulted in isolation of endogenous IKK- α (Fig. 3b) and immunoprecipitation of endogenous IKK- α co-precipitated HA–IKK γ (Fig. 3c). Because of a lack of high-affinity antibodies for IKK- β , we were unable to study the

interaction between HA–IKK γ and endogenous IKK- β . The interaction between IKK- γ and IKK- α was not altered by cytokines (Figs 1c and 3b).

As IKK- α and IKK- β form very stable heterodimers¹⁶, these results do not indicate whether IKK- γ binds directly to IKK- α , IKK- β or both proteins. We therefore studied interactions of IKK- γ with IKK- α or IKK- β using purified recombinant proteins (Fig. 3d). We detected direct and stable binding of IKK- γ to IKK- β but not to IKK- α .

Immunoprecipitation of HA–IKK γ from transiently transfected cells resulted in isolation of an I κ B kinase activity that was stimulated by either TNF or IL-1 (Fig. 4a). These results were identical to those obtained when anti-IKK α was used to isolate the IKK complex. This similarity is due to efficient interaction between transiently expressed HA–IKK γ and other IKK components. Gel-filtration analysis indicated that HA–IKK γ was incorporated into the large, 900K IKK complex, precisely co-eluting with IKK- α (Fig. 4b).

We first studied the involvement of IKK- γ in IKK activation through expression of IKK- γ antisense RNA. Whereas co-transfection of an IKK- γ sense vector had no effect on I κ B kinase activity associated with either IKK- α or IKK- β (see below), the level of TNF-induced I κ B kinase activity associated with either IKK subunit decreased upon co-transfection with the IKK- γ antisense vector (Fig. 5a). Antisense IKK- γ reduced the expression of HA–IKK γ , but had no effect on the expression of either HA–IKK α or HA–IKK β (Fig. 5b). A cDNA that can complement IKK activity in two variant cell lines that are defective in NF- κ B activation has recently been identified¹⁷. The product of this cDNA, named NF- κ B essential modulator (NEMO), seemed to be the mouse homologue of IKK- γ .

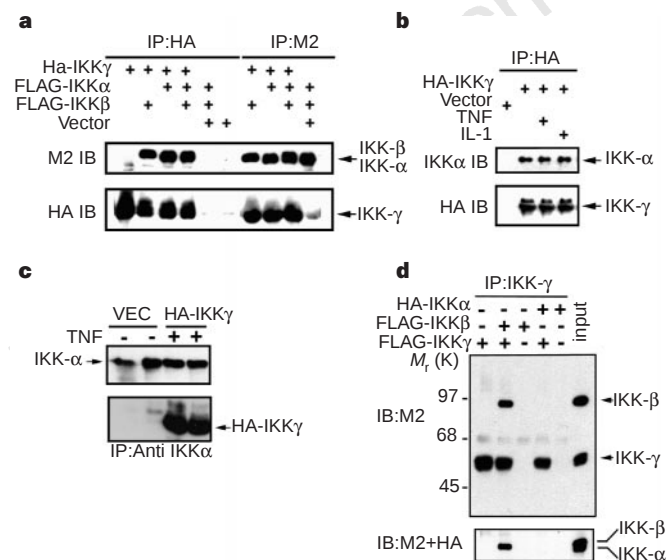


Figure 3 IKK- γ interacts physically with IKK- α / β . **a**, HA–IKK γ , Flag–IKK α , Flag–IKK β or ‘empty’ expression vectors were transiently transfected into 293 cells as indicated. After 24 h the cells were lysed. Part of each lysate was precipitated with anti-HA antibody and another part with anti-Flag antibody (M2). The levels of Flag–IKK α , Flag–IKK β and HA–IKK γ were determined by immunoblotting. **b**, HA–IKK γ or empty vectors were transfected into HeLa cells. After 24 h the cells were left untreated or incubated with either TNF or IL-1, lysed, immunoprecipitated with anti-HA antibody and immunoblotted with anti-IKK α and anti-HA antibody. **c**, HA–IKK γ or empty (VEC) vectors were transfected into 293 cells, which were treated and processed as in **b**. **d**, HA–IKK α and Flag–IKK β were expressed in Sf9 cells using baculovirus vectors and were purified¹⁶. They were incubated with or without purified recombinant Flag–IKK γ . The proteins were immunoprecipitated with anti-IKK γ (anti-NEMO)¹⁷ antibody and immunoblotted with anti-HA and anti-Flag (M2) antibody. IP, immunoprecipitation; IB, immunoblot.

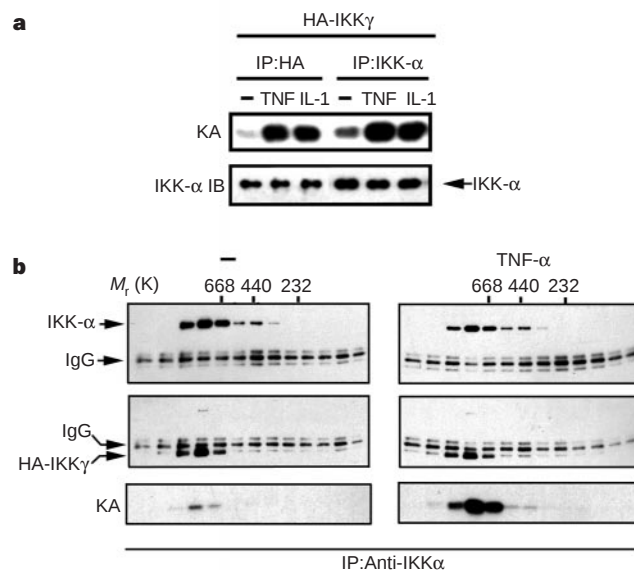


Figure 4 IKK- γ is a component of the I κ B kinase complex. **a**, HeLa cells were transiently transfected with HA–IKK γ . After 24 h cells were or were not treated with TNF or IL-1. Part of each lysate was immunoprecipitated (IP) with anti-HA antibody and another part with anti-IKK α antibody, and I κ B kinase activity (KA) was determined as described¹. Levels of endogenous IKK- α were determined by immunoblotting (IB) with anti-IKK α antibody. **b**, Extracts of unstimulated or TNF-treated 293 cells that were transfected with an HA–IKK γ vector were fractionated on a Superose 6 column. Fractions were immunoprecipitated with anti-IKK α antibody. Immune complex kinase assays (KA) and immunoblotting of IKK- α and HA–IKK γ were done with IKK- α and HA antibodies.

Using anti-NEMO antibodies (a gift from A. Israel), we confirmed this identity and found that transient expression of antisense IKK- γ RNA reduced expression of endogenous IKK- γ but not of IKK- α (Fig. 5c) or IKK- β (data not shown).

Antisense IKK- γ also reduced the extent of IKK activation by IL-1 or transiently transfected MEKK-1 or NIK vectors, although the inhibition of the response to overexpressed NIK was considerably less than that of the other responses (Fig. 5d). Co-transfection of antisense vectors for the kinases JNKK-1 and MKK-3 had no effect on IKK activity¹ and antisense IKK- γ did not inhibit the activation of the kinase p38^{MAPK} by either TNF or IL-1 (Fig. 6e). Transient expression of antisense IKK- γ prevented TNF-induced nuclear entry of the RelA (p65) subunit of NF- κ B (data not shown). We also established stably transfected pools of 293 cells containing the antisense IKK- γ expression vector. Cells in these pools expressed less IKK- γ (Fig. 7b) and, compared with the parental cells, exhibited less TNF-induced I κ B- α phosphorylation and degradation (Fig. 7a) and NF- κ B activation (Fig. 7c). This inhibitory effect was specific as IKK- α expression was not decreased (Fig. 7a) and the DNA-binding activity of the constitutive transcription factor NF-1 was actually increased in cells transfected with antisense IKK- γ (Fig. 7c). In addition, the antisense-IKK- γ -transfected cells exhibited normal activation of the kinases JNK and p38^{MAPK} (Fig. 7d).

To further study the function of IKK- γ and gain clues as to how it is involved in regulation of IKK activity, we constructed N- and C-terminal deletion mutants (Fig. 6a) and studied them for possible dominant inhibitory activity. Expression of Δ N-IKK γ (134–419) with Flag-IKK β had only a marginal effect on basal IKK activity and its response to TNF. However, expression of Δ C-IKK γ (1–300) inhibited activation of IKK by TNF but not basal kinase activity (Fig. 6b). Both Δ N-IKK γ (134–419) and Δ C-IKK γ (1–300) inter-

acted with IKK- α/β in cells (Fig. 6c), but only full-length IKK γ and Δ N-IKK γ (134–419) were able to co-precipitate IKK activity stimulated by TNF, IL-1, MEKK-1 or NIK (data not shown). Cross-linking experiments using recombinant proteins indicated that IKK- γ can form dimers and trimers and that the C-terminal truncation had no effect on this activity, although the N-terminal truncation may have reduced the efficiency of trimerization. Neither full-length IKK- γ nor its truncation mutants inhibited activation of p38^{MAPK} by TNF or IL-1 (Fig. 6e).

Previous studies showed that IKK contains two catalytic subunits, IKK- α and IKK- β ^{2,4,5}. The identity of the remaining IKK subunits and their function were unknown until now. We have purified IKK to homogeneity and found that, in addition to IKK- α/β , it contains two other major polypeptides, IKK- γ 1 and IKK- γ 2, which are derived from the same transcript. The presence of leucine zippers and coiled-coil motifs indicates that IKK- γ can be involved in homotypic and heterotypic interactions. Indeed, recombinant IKK- γ forms dimers and trimers and can directly interact with IKK- β . As IKK- α and IKK- β form stable homodimers and heterodimers even in the absence of IKK- γ ¹⁶, the core IKK complex might consist of an IKK α -IKK β heterodimer associated with an IKK- γ dimer or trimer.

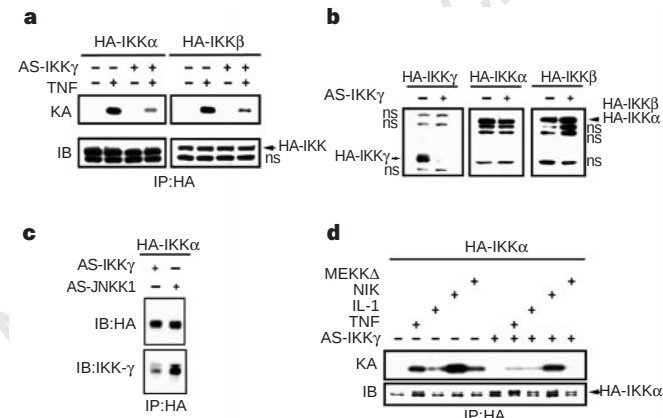


Figure 5 IKK- γ is an essential component of the I κ B kinase. **a**, HA-IKK α and HA-IKK β vectors were transfected into HeLa cells together with either 'empty' or antisense (AS)-IKK- γ vectors. After 24 h cells were or were not treated with TNF and were lysed. Lysates were precipitated with anti-HA antibody and IKK activity was determined by immune complex kinase assays (KA). Expression of HA-IKK α and HA-IKK β was determined by immunoblotting. Migration positions of IKK- α and IKK- β and of a nonspecific (ns) band are indicated. **b**, HA-IKK γ , HA-IKK α or HA-IKK β vectors were transfected into 293 cells together with either 'empty' or AS-IKK γ vectors as indicated. After 24 h the cells were lysed, immunoprecipitated with anti-HA antibody and immunoblotted with anti-HA antibody. **c**, HeLa cells were co-transfected with HA-IKK α and either AS-IKK γ or AS-JNKK1 vectors. After 24 h cell lysates were prepared, immunoprecipitated with anti-HA antibody and immunoblotted with anti-HA antibody or anti-NEMO (anti-IKK γ) antibody. **d**, HeLa cells were transfected with HA-IKK α and either 'empty' or AS-IKK γ vectors, together with NIK or MEKK-1 catalytic domain (MEKK Δ) expression vectors, as indicated. After 24 h cells were or were not treated with TNF or IL-1, and were lysed and immunoprecipitated with anti-HA antibody. IKK activity was determined as above. IP, immunoprecipitation; IB, immunoblot.

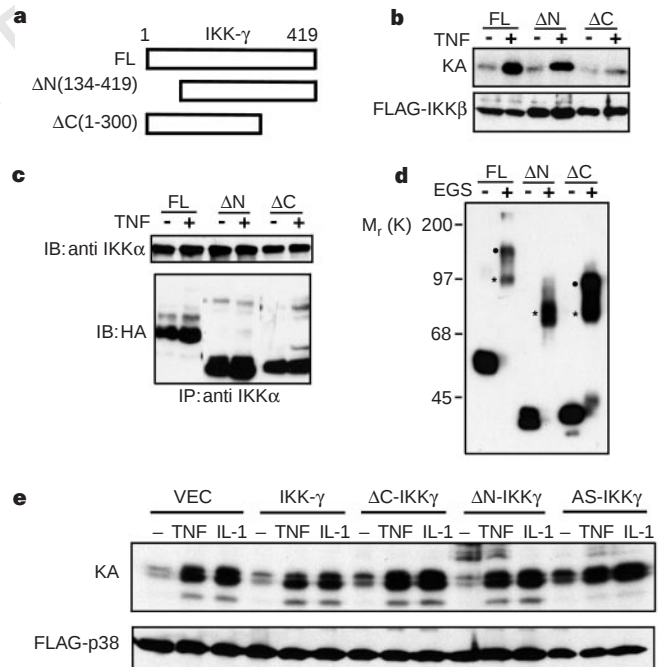


Figure 6 A C-terminal IKK- γ deletion mutant is a dominant-negative inhibitor of IKK activation. **a**, Full-length (FL) IKK- γ and its deletion mutants. **b**, Flag-IKK β was transfected together with wild-type and truncated IKK- γ expression vectors into HeLa cells. After 24 h cells were or were not incubated with TNF and were then lysed. The lysates were immunoprecipitated with anti-Flag antibody (M2), and IKK kinase activity (KA) and Flag-IKK β expression were determined. **c**, 293 cells were transfected with the different HA-IKK γ vectors and treated as above. Lysates were immunoprecipitated with anti-IKK α antibody and immunoblotted (IB) with anti-IKK α and anti-HA antibodies. **d**, Recombinant full-length IKK- γ and its truncation mutants were expressed in *E. coli*, purified and incubated with or without the crosslinking agent ethylene glycolbis(succinimidylsuccinate) (EGS). The proteins were separated by SDS-PAGE and visualized by immunoblotting. The asterisks and dots mark the dimer and trimer bands, respectively. **e**, Flag-p38 vector was transfected together with either 'empty' vector (VEC) or IKK- γ sense and antisense (AS) expression vectors. After 24 h the cells were either left untreated or were incubated with TNF or IL-1. Lysates were prepared, and p38 activity and expression were determined by immune complex kinase assay (KA) and immunoblotting, respectively.

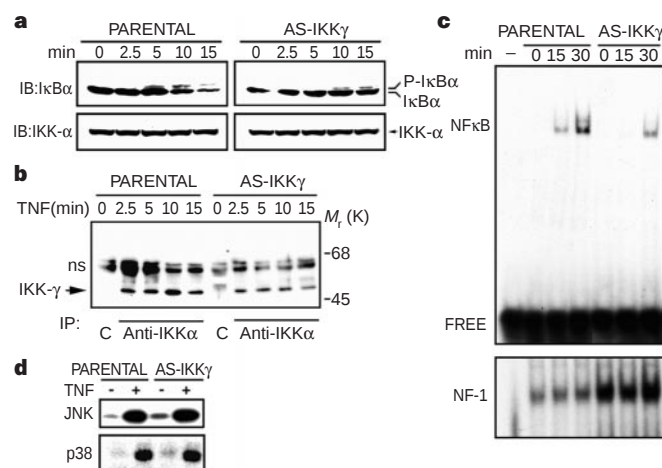


Figure 7 Reduced IKK- γ expression interferes with I κ B- α phosphorylation and degradation and NF- κ B activation. **a**, Pools of 293 cells stably transfected with AS-IKK γ vector and parental mock-transfected cells were stimulated with TNF for the indicated times, after which cells were collected and lysed. Lysates were tested by immunoblotting for I κ B- α degradation and endogenous IKK- α levels. Basal and phosphorylated (P-I κ B α) forms of I κ B- α are indicated. **b**, IKK- α immune complexes were isolated and immunoblotted with anti-NEMO (anti-IKK γ) antibody. **c**, immunoprecipitation with control antibody; ns, nonspecific band. **d**, Parental cells or pools of 293 cells stably transfected with the AS-IKK γ vector were incubated with TNF for the indicated times, after which nuclear extracts were prepared. The levels of DNA-binding activities of NF- κ B and NF-1 were determined by electrophoretic mobility shift assay¹⁸. **d**, JNK and p38 activities were determined by immune complex kinase assay. IP, immunoprecipitation; IB, immunoblot.

Although the exact biochemical function of IKK- γ is yet to be determined, it is an essential component of IKK. Reduced IKK- γ expression results in decreased IKK activation (Figs 5 and 7) and its complete absence abolishes IKK and NF- κ B activation altogether¹⁷. Although IKK activity is absolutely dependent on IKK- α / β dimerization⁵, IKK- γ is unlikely to function as a chaperone or a co-factor that stabilizes IKK α -IKK β dimers. The ability of the C-terminally truncated IKK- γ mutant to inhibit IKK activation by upstream stimuli, while having only a small effect on basal kinase activity, indicates that the major function of IKK- γ may be to connect the IKK complex to upstream activators. This function is likely to be mediated by the C-terminal region of IKK- γ , while its central region probably interacts with IKK- β . Although, *in vitro*, IKK- γ stably interacts with IKK- β and not with IKK- α , it is likely that once recruited into the complex IKK- γ also interacts with IKK- α .

Methods

Purification and cloning of IKK- γ . The IKK complex was purified from HeLa and Jurkat cells as described¹ except for substituting affinity chromatography on an I κ B α (1–54) column with affinity chromatography on immobilized monoclonal anti-IKK α antibody (B78–743) (this antibody is available from PharMingen), generated against full-length recombinant IKK- α . Purified antibody (0.5 mg) was coupled to 0.3 ml CNBr-activated Sepharose 4B (Pharmacia). Active IKK fractions, after gel filtration on Superose 6, were pooled (1.6 ml) and applied batchwise to the immunoaffinity resin (0.1 ml). The mixture was rotated at 4°C for 4 h and then centrifuged. The beads were washed with 20 ml buffer A¹ containing 500 mM NaCl and 1% Triton X-100 and then with 5 ml buffer A containing 2 M urea. Bound protein was eluted with 0.5% SDS and separated by SDS-PAGE. Bands identified as IKK- γ were transferred to a PVDF membrane, stained with colloidal blue and digested with Lys-C. The IKK- γ 1 band was also obtained in pure form after gel filtration and

SDS-PAGE without requiring the immunoaffinity step. This material was transferred to a PVDF membrane and digested with Lys-C. A total of 15 peptide sequences derived from the 52K (IKK- γ 1) and 50K (IKK- γ 2) forms was obtained (by T. W. Thannhauser, data not shown). All of these sequences are contained within the IKK- γ ORF (Fig. 2). The peptide maps generated by Lys-C digestion of IKK- γ 1 and IKK- γ 2 are very similar (data not shown). These sequence data were used to search Genebank and two overlapping human ESTs were found (accession numbers AA133061 and R56495). Primers derived from these ESTs were used to generate a probe to isolate a clone containing the complete IKK- γ ORF from a HeLa cDNA library. The complete IKK- γ nucleotide sequence is available under accession number AF074382.

Kinase assays, immunoprecipitation, immunoblotting, immunofluorescence and gel-shift assay. Kinase assays, immunoprecipitation, immunoblotting and indirect immunofluorescence experiments were done as described¹⁵. The monoclonal anti-IKK α antibody was used for immunoblotting and immunoprecipitation. This antibody does not crossreact with IKK- β (data not shown). TNF- α and IL-1 were used at 20 ng ml⁻¹ and 10 ng ml⁻¹, respectively. Induction times were 10 min except where indicated otherwise (in figure legends). The probes used in the gel-shift assay correspond to the consensus κ B (5'-AGTTGAGGGGACTTCCAGGC-3') and NF-1 (5'-TTGGATTGAAGCCAATATGATA-3') sites.

Plasmids, cell culture, transfections and ³⁵S labelling. The expression vectors were constructed using standard recombinant DNA procedures. The β -actin promoter was used to drive expression of all sense IKK- γ constructs. Antisense IKK- γ constructs were made in vector pcDNA3.1/MyC-His (Invitrogen). Cell culture and transfections were as described¹ except that Lipofectamine Plus (Gibco) was used. IKK mutants were generated as described⁵. ³⁵S labelling was carried out using Pro-Mix (Amersham).

Expression and purification of recombinant proteins. Recombinant HA-IKK α and Flag-IKK β were expressed in Sf9 cells using recombinant baculovirus vectors and were purified to homogeneity¹⁶. Recombinant hexahistidine-tagged IKK- γ proteins (full-length and truncation mutants) were expressed in *Escherichia coli* and purified by nickel-affinity chromatography.

Received 4 June; accepted 13 August 1998.

- DiDonato, J. A., Hayakawa, M., Rothwarf, D. M., Zandi, E. & Karin, M. A cytokine-responsive I κ B kinase that activates the transcription factor NF- κ B. *Nature* **388**, 548–554 (1997).
- Mercurio, F. et al. IKK-1 and IKK-2: cytokine-activated I κ B kinases essential for NF- κ B activation. *Science* **278**, 860–866 (1997).
- Régnier, C. H. et al. Identification and characterization of an I κ B kinase. *Cell* **90**, 373–383 (1997).
- Woronicz, J. D., Gao, X., Cao, Z., Rothe, M. & Goeddel, D. V. I κ B kinase- β : NF- κ B activation and complex formation with I κ B kinase- α and NIK. *Science* **278**, 866–869 (1997).
- Zandi, E., Rothwarf, D., Delhase, M., Hayakawa, M. & Karin, M. The I κ B kinase complex (IKK) contains two kinase subunits, IKK α and IKK β , necessary for NF- κ B activation. *Cell* **91**, 243–252 (1997).
- Verma, I. M., Stevenson, J. K., Schwarz, E. M., Van Antwerp, D. & Miyamoto, S. Rel/NF- κ B/I κ B family: intimate tales of association and dissociation. *Genes Dev.* **9**, 2723–2735 (1995).
- Baeuerle, P. A. & Baltimore, D. NF- κ B: ten years after. *Cell* **87**, 13–20 (1996).
- Baeuerle, P. A. & Henkel, T. Function and activation of NF- κ B in the immune system. *Annu. Rev. Immunol.* **12**, 141–179 (1994).
- Barnes, P. J. & Karin, M. Nuclear factor- κ B—a pivotal transcription factor in chronic inflammatory diseases. *New Engl. J. Med.* **336**, 1066–1071 (1997).
- Ling, L., Cao, Z. & Goeddel, D. V. NF- κ B-inducing kinase activates IKK- α by phosphorylation of Ser-176. *Proc. Natl Acad. Sci. USA* **95**, 2791–2797 (1998).
- Nakano, H. et al. Differential regulation of I κ B kinase α and β by two upstream kinases, NF- κ B-inducing kinase and mitogen-activated protein kinase/ERK kinase kinase-1. *Proc. Natl Acad. Sci. USA* **95**, 3537–3542 (1998).
- Karin, M. & Delhase, M. JNK or IKK, AP-1 or NF- κ B, which are the targets for MEKK1 action? *Proc. Natl Acad. Sci. USA* **95**, 9067–9069 (1998).
- Berger, B. et al. Predicting coiled coils by use of pairwise residue correlations. *Proc. Natl Acad. Sci. USA* **92**, 8259–8263 (1995).
- Rost, B. PHD: predicting one-dimensional protein structure by profile based neural network. *Methods Enzymol.* **266**, 525–539 (1996).
- Frishman, D. & Argos, P. Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence. *Protein Eng.* **9**, 133–142 (1996).
- Zandi, E., Chen, Y. & Karin, M. Direct phosphorylation of I κ B by IKK α and IKK β : discrimination between free and NF- κ B-bound substrate. *Science* **281**, 1360–1363 (1998).
- Yamaoka, S. et al. Complementation cloning of NEMO, a component of the I κ B kinase complex essential for NF- κ B activation. *Cell* **93**, 1231–1240 (1998).
- DiDonato, J. A., Mercurio, F. & Karin, M. Phosphorylation of I κ B precedes but is not sufficient for its dissociation from NF- κ B. *Mol. Cell. Biol.* **15**, 1302–1322 (1995).

Acknowledgements. We thank T. Thannhauser for help with peptide sequencing, A. Israel for the NEMO antibody and sharing of unpublished information, F. Mercurio for the IKK- β antibody, G. Cadwell and J. Lewis for technical assistance and B. Thompson for manuscript preparation. D.M.R., E.Z. and G.N. were supported for postdoctoral fellowships by the NIH, the Leukemia Society of America and the Damon–Runyon Walter–Winchell Cancer Research Fund, respectively. This work was supported by grants from the NIH.

Correspondence and requests for materials should be addressed to M.K. (e-mail: karinoffice@ucsd.edu).

Signal-specific co-activator domain requirements for Pit-1 activation

Lan Xu^{††}, Robert M. Lavinsky^{‡‡}, Jeremy S. Dasen^{††}, Sarah E. Flynn[‡], Eileen M. McInerney[‡], Tina-Marie Mullen^{§§}, Thorsten Heinzel[‡], Daniel Szeto^{††}, Edward Korzus[‡], Riki Kurokawa^{§§}, Aneel K. Aggarwal[‡], David W. Rose[§], Christopher K. Glass^{§§} & Michael G. Rosenfeld^{‡‡}

^{*} Howard Hughes Medical Institute, [†] Biomedical Sciences Graduate Program, [‡] Department of Biology Graduate Program, [§] Whittier Diabetes Program, ^{||} Cellular and Molecular Medicine, and [¶] Department and School of Medicine, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093-0648, USA

[☆] Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029-6574, USA

POU-domain proteins, such as the pituitary-specific factor Pit-1, are members of the homeodomain family of proteins which are important in development and homeostasis, acting constitutively or in response to signal-transduction pathways to either repress or activate the expression of specific genes¹. Here we show that whereas homeodomain-containing repressors such as Rpx² seem

to recruit only a co-repressor complex, the activity of Pit-1 (ref. 3) is determined by a regulated balance between a co-repressor complex that contains N-CoR/SMRT^{4,5}, mSin3A/B⁶⁻⁸ and histone deacetylases⁶⁻⁸, and a co-activator complex that includes the CREB-binding protein (CBP)⁹ and p/CAF¹⁰. Activation of Pit-1 by cyclic AMP or growth factors depends on distinct amino- and carboxy-terminal domains of CBP, respectively. Furthermore, the histone acetyltransferase functions of CBP^{11,12} or p/CAF¹⁰ are required for Pit-1 function that is stimulated by cyclic AMP or growth factors, respectively. These data show that there is a switch in specific requirements for histone acetyltransferases and CBP domains in mediating the effects of different signal-transduction pathways on specific DNA-bound transcription factors.

A potential connection between the nuclear receptor co-repressor (N-CoR) and homeodomain-containing transcription factors was indicated by screening of λ gt11 complementary DNA libraries using the radio-labelled, bacterially expressed repression domain III (amino acids 970–1,502) of N-CoR. Unexpectedly, one-third of the sequenced positive isolates encompassed the homeodomain or POU domain of several transcriptional activators. Therefore, we tested whether Pit-1, a tissue-specific POU homeodomain factor that is important in the differentiation of three pituitary cell types³, could associate with N-CoR, and whether such an interaction might modulate Pit-1 function. When acting alone, Pit-1 weakly activated the minimal promoters under the control of a multimerized cognate Pit-1 response element, which is derived from the prolactin Pit-1 1P site (Fig. 1a). Blocking N-CoR, mSin3A/B or histone deacetylase

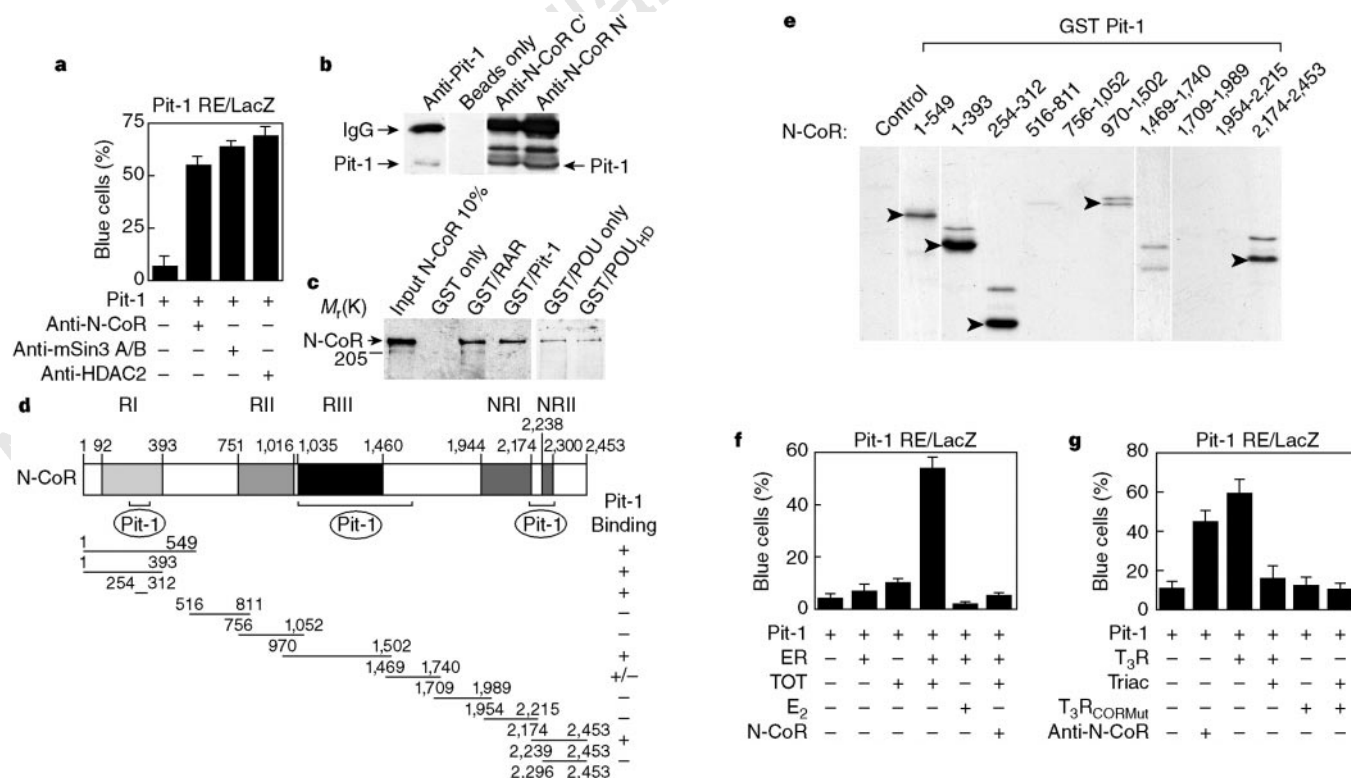


Figure 1 Role of N-CoR repression in Pit-1 function. **a**, Single-cell microinjection assay showing effects of anti-N-CoR, anti-mSin3 A/B, or anti-HDAC²³ IgG on Pit-1-mediated activation of a Pit-1-dependent reporter (Pit-1 RE/LacZ). **b**, Immunoprecipitation with whole-cell extract prepared from 293 cells transfected with Pit-1 expression vector, using specific antibodies against Pit-1 and the N or C terminus of N-CoR (N-terminal amino acids 1–132, C-terminal amino acids 2,239–2,456). The western blot was detected with anti-Pit-1 antibody. **c**, GST pull-down from extracts of 293 cells transfected with cytomegalovirus (CMV)-driven Flag-tagged N-CoR. GST fusions containing the retinoic acid receptor, Pit-1, the POU domain of Pit-1 or POU_{HD} of Pit-1 were used. The western blot was detected with anti-Flag

antibody (Kodak). **d**, Map of N-CoR showing locations of three independent Pit-1 interaction domains. RI, RII, RIII: repression domains I, II and III (refs 4, 14); NRI, NR II: nuclear-receptor-binding domains I and II (refs 4, 14). **e**, Interaction assay using GST-Pit-1 and ³⁵S-labelled N-CoR fragments (TNT, Promega). **f**, The ability of TOT-occupied oestrogen receptor (ER) to 'transactivate' Pit-1 was studied using the reporter as in **a**. ER exhibited TOT-dependent (10⁻⁷ M), but not 17- β -oestradiol (E₂)-dependent (10⁻⁷ M), activation of the Pit-1-responsive reporter. **g**, The thyroid-hormone receptor CoR box mutant (T₃R CoRmut) was used to determine whether N-CoR binding by the thyroid-hormone receptor (T₃R) was responsible for the ability of unliganded T₃R to activate Pit-1 'in trans'.

HDAC2 (ref. 13) functions by antibodies in the microinjection assay resulted in markedly increased Pit-1 activity (Fig. 1a). The same antibodies had no effects on control reporters (Fig. 2b, and data not shown). Immunoprecipitation with N-CoR antibodies resulted in co-precipitation of Pit-1 (Fig. 1b); and bacterially produced full-length Pit-1 or the POU domain alone were able to specifically pull down Flag-epitope-tagged N-CoR expressed in 293 cells (Fig. 1c). The Pit-1 POU domain interacted specifically with N-CoR repression regions I and III (refs 4, 14), and a C-terminal region of N-CoR that is N-terminal to the primary interaction domain of the nuclear receptor^{4,14} (NR-I; Fig. 1d, e). These results indicate that the N-CoR complex exerts significant inhibitory effects on Pit-1 activity within the cell.

We next investigated whether the N-CoR complex could be 'limiting' in a fashion analogous to that suggested for co-activators in mediating transrepression between nuclear receptor and AP-1 (ref. 15). The observation that 4-*trans*-hydroxytamoxifen (TOT), an antagonist of the oestrogen receptor, but not the agonist 17- β -oestradiol, induces N-CoR interaction with the oestrogen receptor¹⁶ allowed us to examine whether antagonist-bound oestrogen receptor would be able to confer 'transactivation' to Pit-1 by titrating the available N-CoR. Indeed, oestrogen receptor bound to TOT, but not to 17- β -oestradiol, resulted in transactivation of Pit-1, and this effect was abolished by overexpression of N-CoR (Fig. 1f).

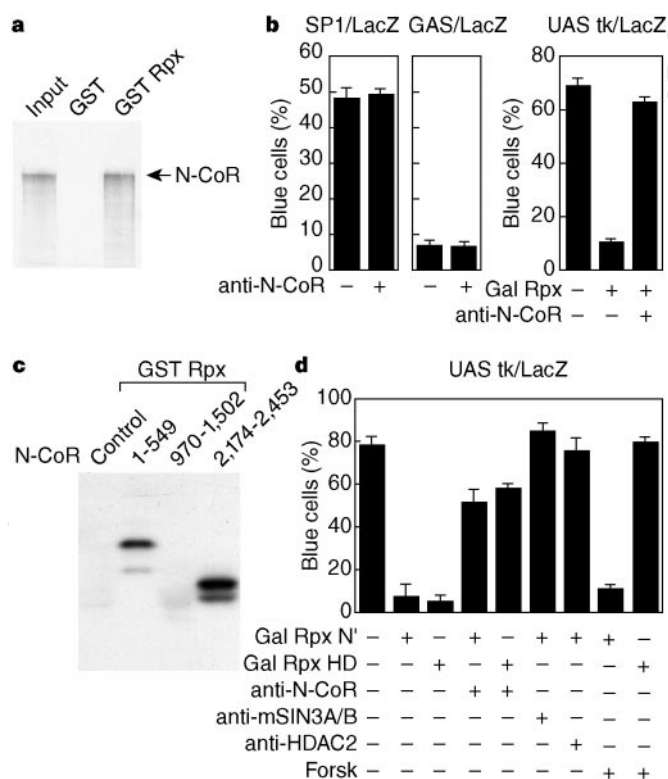


Figure 2 Requirement for N-CoR co-repressor complex in the repression function of Rpx. **a**, GST-Rpx interacted with full-length N-CoR expressed by *in vitro* transcription and translation (TNT, Promega). **b**, Rpx fused to the Gal4 DNA-binding domain was able to repress a high baseline reporter upstream activation sequence (UAS) tk/LacZ²⁶. This repression was largely abolished by injection of anti-N-CoR IgG. Control reporters driven by Sp-1 sites or interferon- γ -responsive (GAS) elements were not affected in the same assay. **c**, Mapping of the Rpx-interaction domain on N-CoR. **d**, Injection of purified IgGs against N-CoR, mSin3 A/B and HDAC2 (ref. 13) blocked the repression function of both the N terminus and homeodomain of Rpx. Forskolin treatment reduced the repression function that was mediated by the homeodomain of Rpx. The reporter system is as described in (b).

Similarly, unliganded, wild-type thyroid-hormone receptor, which binds to N-CoR, caused strong transactivation of Pit-1, and this was reversed by dismissal of N-CoR from the thyroid-hormone receptor after the addition of ligand (Fig. 1g). This transactivation effect was not observed with an unliganded thyroid-hormone receptor CoR box mutant that is unable to bind N-CoR (Fig. 1g)¹⁴. These data support a model in which N-CoR, under certain conditions, can be partitioned between multiple transcription factors, providing an additional level of regulation (see below).

Consistent with its role as a co-repressor, N-CoR could interact with homeodomain repressors such as Rpx² (Fig. 2a, c, and data not shown). Repression mediated by either the N terminus or homeodomain of Rpx required N-CoR, mSin3A/B and HDAC2, as shown in the microinjection assay (Fig. 2b, d). Thus, the repression function of Rpx seemed to result from the cooperative recruitment of N-CoR to both the N terminus and the homeodomain. Repression that was mediated by the Rpx homeodomain, but not its N terminus, could be relieved by forskolin treatment, which increased the level of intracellular cAMP (Fig. 2d).

Because CBP is involved in activities of many families of transcription factors, we investigated whether the activation function of Pit-1 might be linked to recruitment of the CBP co-activator complex. CBP can interact with Pit-1, both by immunoprecipitation with Pit-1 antibody using cells transfected with Pit-1, and by glutathione S-transferase (GST) pull-down assay (Fig. 3a). The most robust interaction mapped primarily to the N terminus of CBP (amino acids 271–450), and represents a new interaction interface on CBP (Fig. 3a). Weaker interaction was also detectable between Pit-1 and a region of CBP that include the C/H3-E1A-binding domain¹⁷. Mapping indicated that this interaction probably involved both the POU_S and the POU_{HD} regions of Pit-1 (Fig. 3b). The role of CBP in Pit-1 function was tested initially by transient co-transfection assays, which showed that overexpression of CBP enhanced Pit-1 action markedly (Fig. 3c).

Analogous to the ligand-induced displacement of the N-CoR co-repressor complex by the CBP co-activator complex in nuclear-receptor activation, we found that N-CoR and CBP can compete directly for binding to Pit-1. By using a DNA-dependent protein-protein interaction assay¹⁸ to test for possible direct competition between N-CoR and CBP for association with Pit-1, we found that the Pit-1-interaction domain of CBP (amino acids 1–450) did indeed compete with N-CoR for association with Pit-1 bound to its cognate DNA site (Fig. 3d). GST alone had no effect (Fig. 3e). Furthermore, anti-CBP IgG reversed the activation of Pit-1 caused by anti-N-CoR IgG (Fig. 3f), which shows directly the functional antagonism between the co-activator and co-repressor complexes. Extra evidence supporting the model that Pit-1 activity is determined by the relative strength of the interactions between N-CoR and CBP was obtained by systematically mutating surface residues in the Pit-1 homeodomain, on the basis of the crystal structure of Pit-1 on a cognate DNA site¹⁹. We found a single point mutation (E254A, in the linker preceding helix 3) that inhibited Pit-1 activity (Fig. 3g, h) without disrupting its ability to bind to DNA (data not shown) or to CBP (Fig. 3i), while exhibiting increased affinity for N-CoR (Fig. 3i).

Increasing intracellular cAMP levels, or activating mitogen-activated protein (MAP) kinases by growth factors such as epidermal growth factor (EGF) or insulin, stimulated the prolactin and Pit-1 promoters^{20–24}. This effect was actually mapped to Pit-1-binding sites themselves, suggesting a connection between the activation of signal-transduction pathways and Pit-1 function^{20,21,24}. In addition, a Pit-1/Ets compound site was defined on the prolactin 3P site and may mediate stimulation of prolactin expression by growth factors²². Here we used the reporter driven by the prolactin 1P element² (Pit-1 RE/Lac Z), which allowed us to examine the impact of signalling pathways on Pit-1 itself. Treating the cells with forskolin or 8-Br cAMP to elevate intracellular cAMP levels, or

activating MAP-kinase pathways by growth factors such as EGF or insulin, resulted in increased Pit-1 activity (Fig. 4a, b). A Pit-1 protein (Pit_{mut}) with mutations in the protein kinase A (PKA) site within the homeodomain²⁴ (T219A/T220A) was fully competent to respond to elevated cAMP level or to insulin/EGF treatment (Fig. 4a), indicating that regulation was not conferred by phosphoryla-

tion of Pit-1 itself, which is consistent with previous reports²³. To further confirm that the actions of cAMP and growth factors were not due to cryptic factors binding to the Pit-1 site, we studied a Gal4–Pit-1 fusion protein. The activity of Gal4–Pit-1 could still be stimulated in response to both forskolin or growth factors, and this activation depends on CBP and p/CAF (Fig. 4b). The stimulation of

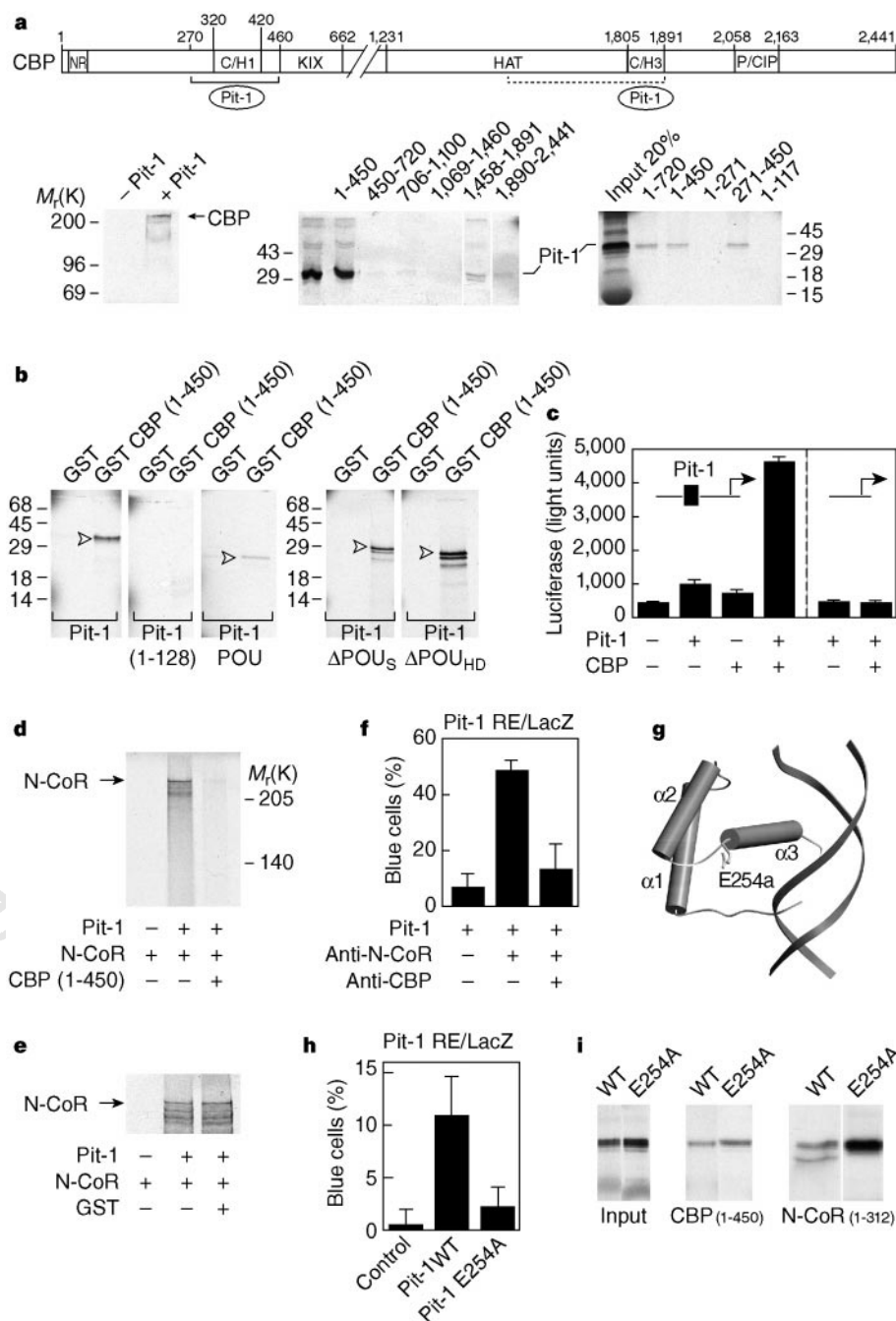


Figure 3 CBP is a required co-activator of Pit-1. **a**, The C/H1 domain of CBP mediates interaction with Pit-1. Whole-cell extracts prepared from 293 cells with or without transfected Pit-1 plasmid were mixed with extracts containing Flag-tagged CBP. Immunoprecipitation was done using anti-Pit-1 antisera and the western blot was detected with anti-Flag antibody (Kodak). GST-fusion proteins of CBP were tested for interaction with ³⁵S-labelled full-length Pit-1. A strong interaction domain was mapped to amino acids 271–450 of CBP, whereas a second weaker interaction domain was mapped to amino acids 1,458–1,891. **b**, Although the combination of POU_S and POU_{HD} (POU) is sufficient for interaction with CBP1–450, deletion of either POU_S or POU_{HD} alone did not eliminate the interaction. **c**, Co-transfection of Pit-1 and CBP led to synergy on a reporter driven by a Pit-1-responsive element. **d**, An avidin/biotin/DNA-dependent protein-

protein interaction assay¹⁸ was done using a biotinylated Pit-1 site (Prl-1P) and bacterially expressed His₆–Pit-1 holoprotein (1 μg), GST–CBP (1–450) (5 μg) and ³⁵S-labelled full-length N-CoR produced by TNT (Promega). GST–CBP1–450 effectively prevented N-CoR from binding to Pit-1. **e**, In the control experiment, GST alone did not block N-CoR binding to Pit-1. **f**, Antagonizing roles of N-CoR and CBP in regulating Pit-1 activity. Blockade of N-CoR repression function led to activation of the Pit-1 response element (RE)/LacZ reporter, and this activation was eliminated with anti-CBP IgG. **g**, Pit-1–DNA crystal structure¹⁹ and position of the E254A mutation in Pit-1. **h**, The E254A mutant had much lower activity on the Pit-1-RE reporter. **i**, The E254A mutant had similar affinity for CBP (1–450) compared to the wild type, but had significantly higher affinity for N-CoR (1–312). WT, wild-type.

Pit-1 by forskolin, or by 8-Br cAMP, did not require SRC-1/NCoA-1 (ref. 25) or p/CIP²⁶, which are crucial for the function of nuclear receptors²⁶ (Fig. 4c).

As phosphorylation of Pit-1 does not appear to account for regulation of Pit-1 activity for forskolin, we considered the possibility that the required components of the CBP co-activator complex were altered by modulation by PKA. Thus, we investigated which functional domain of CBP is critical for the effect of cAMP on Pit-1 activity. Co-injection of expression plasmids that encode CBP holoprotein or CBP mutants with deletion of the CREB binding domain (Δ KIX)⁹, the bromodomain (Δ Br) or the N-terminus (Δ N), all reversed the block of Pit-1 function by anti-CBP IgG, whereas CBP mutants lacking the histone acetyltransferase (HAT) function (Δ HAT)²⁷, the C/H3 domain (Δ C/H3) or the C terminus (CBP1–1,891) failed to do so (Fig. 4d). In contrast, the p/CIP-binding domain of CBP (Δ p/CIP)^{15,26} was not vital for cAMP-stimulated Pit-1 activity (Fig. 4d). All CBP constructs were expressed at roughly equivalent levels; more importantly, their functional integrity was confirmed by their activities in other signalling pathways (Fig. 5c, and data not shown). Mutation of the only perfect PKA consensus site (S1772) (ref. 9) in CBP resulted in almost complete loss of the ability of CBP to mediate cAMP induction of Pit-1 function (Fig. 4c), which could reflect either a

requirement for phosphorylation of this residue by PKA or the importance of amino acid 1,772 in configuration of this region of CBP.

As both CBP and p/CAF are required for Pit-1 function, it is interesting to determine whether the HAT functions from these two proteins are both required, or if only one of them is involved. Thus, we analysed two point mutations—L1690K/C1691L for CBP, and Y616A/F617A for p/CAF—that abolished their ability to acetylate histones without affecting expression²⁷. The ability of wild-type and HAT mutants to reverse the block of Pit-1 functions by anti-CBP and anti-p/CAF indicated that the HAT activity of CBP was required, whereas the HAT function of p/CAF appeared to exert only a minimal role in the activation of Pit-1 by cAMP (Fig. 4e). Our results show that the HAT function, C/H3 domain, and the C-terminal CBP domains are required for stimulation of Pit-1 activity by PKA.

To determine whether these CBP domains could function in a signal-specific fashion, we evaluated their roles in the stimulation of Pit-1 activity by growth factors. Surprisingly, although both CBP and p/CAF were still required, the crucial regions of CBP required to modulate the effects were different from those required for cAMP stimulation. The N-terminal 468 amino acids were vital, whereas the HAT domain and the C/H3 and the C-terminal regions of CBP

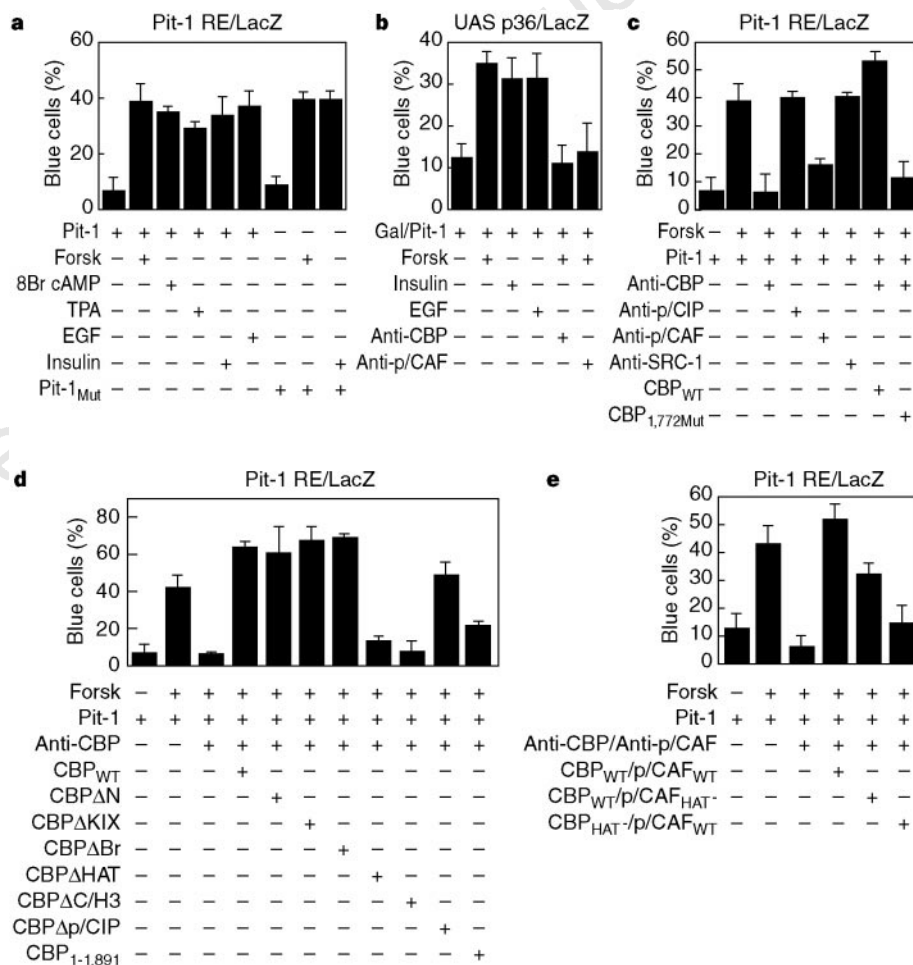


Figure 4 Activation of Pit-1 function by signal-transduction pathways is mediated through the CBP complex. **a**, Pit-1 function is activated by agents stimulating several signal-transduction pathways. Mutations of the PKA sites on Pit-1 (ref. 24) (Pit-1mut: T219A/T220A) did not abolish the ability of both forskolin and insulin to activate Pit-1 function. **b**, The activity of Gal4-Pit-1 can also be stimulated by various signal-transduction pathways. Forskolin-induced activation of Gal4-Pit-1 is dependent on CBP and p/CAF. The reporter contains three copies of UAS sites upstream of the p36 minimal promoter. **c**, Roles of other components of the CBP

complex in Pit-1 function stimulated by forskolin. The consensus PKA site in CBP (S1772) was crucial for induction of Pit-1 activity by forskolin. **d**, Regions of CBP required for forskolin-stimulated Pit-1 function. Indicated CBP mutants were used to rescue the blockade of Pit-1 function by co-injected anti-CBP IgG. CBP Δ N: Δ 2–468; CBP Δ KIX: Δ 451–720; CBP Δ Br: Δ 1,100–1,171; CBP Δ HAT: Δ 1,257–1,413; CBP Δ C/H3: Δ 1,805–1,891; CBP Δ p/CIP: Δ 2,058–2,163. **e**, Requirement of the HAT functions of CBP and p/CAF for Pit-1 function. CBP_{HAT}: L1690K/C1691L; p/CAF_{HAT}: Y616A/F617A.

were no longer required (Fig. 5a). There was no apparent effect of either signal-transduction pathway on levels of expression of any of these mutant CBP proteins, assessed by western blot or by immunohistochemical analysis of injected cells (Fig. 5c, and data not shown). Thus, it appeared that different functional domains of CBP were used by cAMP and growth-factor-induced signalling pathways. Using a mammalian two-hybrid assay with Gal4–CBP1–450 and VP16–Pit-1, we found an enhanced interaction (from 3 to 15% blue cells) between CBP1–450 with Pit-1 in

response to insulin, but not cAMP, indicating that a component of the insulin effect might be to increase interactions between CBP and Pit-1. These results also indicated that P/CAF may provide the HAT function required for growth-factor-stimulated Pit-1 activity. Indeed, the HAT[−] p/CAF was unable to mediate growth-factor-induced Pit-1 activation (Fig. 5b), in contrast to its role in cAMP stimulation of Pit-1 activity (Fig. 4e). These results indicate use of different HAT activities and distinct domains of CBP for cAMP- or growth-factor-induced Pit-1 activity.

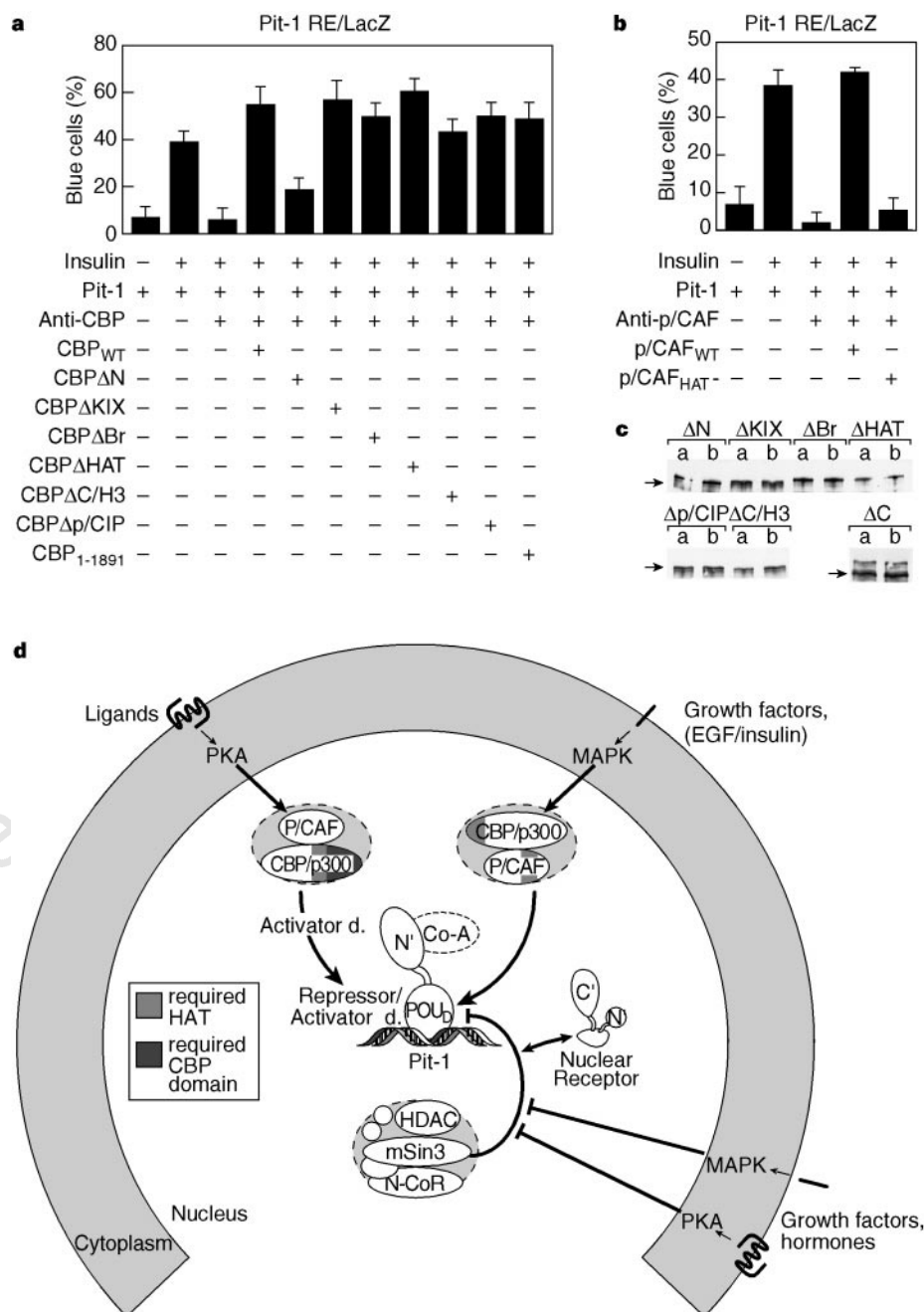


Figure 5 Requirements of HAT functions and domains of CBP for growth-factor-stimulated Pit-1 function. **a**, Regions of CBP required for growth-factor-stimulated Pit-1 function. The CBP mutants are as described in Fig. 4d; note that CBP Δ HAT retains full function in this case. **b**, The HAT function of p/CAF is required for growth-factor-stimulated Pit-1 activity in microinjection assays. **c**, CBP mutants were expressed at comparable levels in transfected cells treated with forskolin (a) or growth factor (b), under conditions similar to those in Fig. 4. In the case of Δ C, antibody against the CBP N terminus was used, and the upper band represents endogenous CBP. The others were detected with anti-Flag antibody. **d**, Model of

modulation of Pit-1 activation by specific signal-transduction pathways. For the PKA-induced Pit-1 activation, the HAT function and C-terminal domain of CBP are required; for growth-factor-induced Pit-1 activation, the HAT function of p/CAF and the N terminus of CBP are required. A putative Pit-1 N-terminal-associated co-activator (Co-A) is proposed to cooperatively recruit the CBP co-activator complex to the POU domain and displace the N-CoR complex. The co-repressor complex, while exerting repression on both Pit-1 and the unliganded nuclear receptor, can be partitioned (two-headed arrow) between the two classes of transcription factors, resulting in 'transactivation'.

Together, these data show that the large family of activating homeodomain transcription factors, such as Pit-1, exert their functions through a balance between the N-CoR co-repressor complexes and the CBP co-activator complexes. Pit-1 activity is regulated by distinct signal-transduction pathways, through mechanisms that do not appear to involve modification of Pit-1 itself, but at least in part through regulation of the recruited co-activator complex. A surprising consequence of these events is that a single transcription factor, Pit-1, actually uses the HAT functions of different proteins, and requires the function of different domains of CBP when activated in response to cAMP or growth-factors (Fig. 5d). These results have implications for the mechanism of integration of signalling events that control complex patterns of gene expression by many classes of transcription factors. □

Methods

Protein-protein interaction assays. All the GST-fusion proteins were expressed and purified as described previously^{4,15}. *In vitro* protein-protein interaction assay, immunoprecipitation, GST pull-down and DNA-dependent protein-protein interaction (ABCD) assays were done as described^{4,6,18}.

Nuclear microinjection, staining and fluorescence microscopy. Microinjection analysis was done as described previously using affinity-purified IgGs^{6,15,26,28}.

Transient transfection. HeLa and 293 cells were maintained in DMEM supplemented with 10% fetal bovine serum. Calcium precipitation mediated transient transfection was done according to standard protocol.

Received 13 March; accepted 29 June 1998.

- Gray, S. & Levine, M. Transcriptional repression in development. *Curr. Opin. Cell Biol.* **8**, 358–364 (1996).
- Hermesz, E., Mackem, S. & Mahon, K. A. Rpx: a novel anterior-restricted homeobox gene progressively activated in the prechordal plate, anterior neural plate and Rathke's pouch of the mouse embryo. *Development* **122**, 41–52 (1996).
- Ingraham, H. A. *et al.* A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. *Cell* **55**, 519–529 (1988).
- Horlein, A. J. *et al.* Ligand-independent repression by the thyroid hormone receptor mediated by a nuclear receptor co-repressor. *Nature* **377**, 397–404 (1995).
- Chen, J. D. & Evans, R. M. A transcriptional co-repressor that interacts with nuclear hormone receptors. *Nature* **377**, 454–457 (1995).
- Heinzel, T. *et al.* A complex containing N-CoR, mSin3 and histone deacetylase mediates transcriptional repression. *Nature* **387**, 43–48 (1997).
- Alland, L. *et al.* Role for N-CoR and histone deacetylase in Sin3-mediated transcriptional repression. *Nature* **387**, 49–55 (1997).
- Laherty, C. D. *et al.* Histone deacetylases associated with the mSin3 corepressor mediate Mad transcriptional repression. *Cell* **89**, 349–356 (1997).
- Chirivia, J. C. *et al.* Phosphorylated CREB binds specifically to the nuclear protein CBP. *Nature* **365**, 855–859 (1993).
- Yang, X.-Y., Ogryzko, V. V., Nishikawa, J., Howard, B. H. & Nakatani, Y. A p300/CBP-associated factor that competes with the adenoviral oncoprotein E1A. *Nature* **382**, 319–324 (1996).
- Ogryzko, V. V., Schiltz, R. L., Russanova, V., Howard, B. H. & Nakatani, Y. The transcriptional coactivator p300 and CBP are histone acetyltransferases. *Cell* **87**, 953–960 (1996).
- Bannister, A. J. & Kouzarides, T. The CBP coactivator is a histone acetyltransferase. *Nature* **384**, 641–643 (1996).
- Yang, W.-M., Inouye, C., Zeng, Y., Bearss, D. & Seto, E. Transcriptional repression by YY1 is mediated by interaction with a mammalian homolog of the yeast global regulator HDAC2/mRPD3. *Proc. Natl. Acad. Sci. USA* **93**, 12845–12850 (1996).
- Kurokawa, R. *et al.* Polarity-specific activities of retinoic acid receptors determined by a co-repressor. *Nature* **377**, 451–454 (1995).
- Kamei, Y. *et al.* A CBP integrator complex mediates transcriptional activation and AP-1 inhibition by nuclear receptors. *Cell* **85**, 403–414 (1996).
- Lavinsky, R. M. *et al.* Diverse signaling pathways modulate nuclear receptor recruitment of N-CoR and SMRT complexes. *Proc. Natl. Acad. Sci. USA* **95**, 2920–2925 (1998).
- Lundblad, J. R., Kwok, R. P., Lurance, M. E., Harter, M. L. & Goodman, R. H. Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP. *Nature* **374**, 85–88 (1995).
- Glass, C. K., Lipkin, S. M., Devary, O. V. & Rosenfeld, M. G. Positive and negative regulation of gene transcription by a retinoic acid-thyroid hormone receptor heterodimer. *Cell* **59**, 697–708 (1989).
- Jacobson, E. M., Li, P., Leon-del-Rio, A., Rosenfeld, M. G. & Aggarwal, A. K. Structure of Pit-1 POU domain bound to DNA as a dimer: unexpected arrangement and flexibility. *Genes Dev.* **11**, 198–212 (1997).
- Elsholtz, H. P., Lew, A. M., Albert, P. R. & Sundmark, V. C. Inhibitory control of prolactin and Pit-1 gene promoters by dopamine. *J. Biol. Chem.* **266**, 22919–22925 (1991).
- Murdoch, G. H., Waterman, M., Evans, R. M. & Rosenfeld, M. G. Molecular mechanism of phorbol ester, thyrotropin-releasing hormone, and growth factor stimulation of prolactin gene transcription. *J. Biol. Chem.* **260**, 11852–11858 (1985).
- Howard, P. W. & Maurer, R. A. A composite Ets/Pit-1 binding site in the prolactin gene can mediate transcriptional responses to multiple signal transduction pathways. *J. Biol. Chem.* **270**, 20930–20936 (1995).
- Okimura, Y., Howard, P. W. & Maurer, R. A. Pit-1 binding sites mediate transcriptional responses to cyclic adenosine 3',5'-monophosphate through a mechanism that does not require inducible

phosphorylation of Pit-1. *Mol. Endocrinol.* **8**, 1559–1565 (1994).

- Kapiloff, M. S., Farkash, Y., Wegner, M. & Rosenfeld, M. G. Variable effects of phosphorylation of Pit-1 dictated by the DNA response elements. *Science* **253**, 786–789 (1991).
- Önate, S. A., Tsai, S. Y., Tsai, M.-J. & O'Malley, B. W. Sequence and characterization of a coactivator for the steroid hormone receptor superfamily. *Science* **270**, 1354–1357 (1995).
- Torchia, J. *et al.* The transcriptional co-activator p/CIP binds CBP and mediates nuclear-receptor function. *Nature* **387**, 677–684 (1997).
- Korzus, E. *et al.* Transcriptional factor-specific requirements for coactivators and their acetyltransferase functions. *Science* **279**, 703–707 (1998).
- Rose, D. W., McCabe, G., Feramisco, J. R. & Adler, M. Expression of c-fos and AP-1 activity in senescent human fibroblasts is not sufficient for DNA synthesis. *J. Cell. Biol.* **119**, 1405–1411 (1992).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank C. Laherty, R. Eisenman, W.-M. Yang and E. Seto for antibodies, C. Nelson for cell culture, L.-M. Philips for assistance, and P. Meyer for her expertise in the preparation of the illustrations. This work was supported by an American Heart Association Predoctoral fellowship (L.X.), the NSF (R.M.L.), the Lucille Markey Fund (J.S.D.), and an NIH predoctoral fellowship (D.S.), a US Army Medical Research Program Award (E.K.), by grants from HHMI and NIH (to M.G.R.), and by an American Diabetes Association Career Development Award (to D.W.R.). C.K.G. is an Established Investigator of the American Heart Association and M.G.R. is an investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to M.G.R. (e-mail: mrosenfeld@ucsd.edu).

errata

Evidence for the shikimate pathway in apicomplexan parasites

Fiona Roberts, Craig W. Roberts, Jennifer J. Johnson, Dennis E. Kyle, Tino Krell, John R. Coggins, Graham H. Coombs, Wilbur K. Milhous, Saul Tzipori, David J. P. Ferguson, Debopam Chakrabarti & Rima McLeod

Nature **393**, 801–805 (1998)

The correct designation of symbols in Figs 1b, d and 3a is as follows. In Fig. 1b, open squares: PABA alone; filled squares: sulphadiazine and PABA; open circles: glyphosate and PABA; filled circles: pyrimethamine and PABA. Figure 1d shows *in vivo* activity against *T. gondii* of glyphosate (100 mg kg⁻¹ day⁻¹) and pyrimethamine (12.5 mg kg⁻¹ day⁻¹), alone or in combination. Open squares, with 50% survival at 30 days: combination of glyphosate and pyrimethamine; filled circles: pyrimethamine alone; open circles: glyphosate alone. Open squares, with 0% survival at 9 days: untreated controls. In Fig. 3a, diamonds: PABA alone; squares: folic acid alone; upright triangles: glyphosate and PABA; inverted triangles: glyphosate and folic acid. □

Phase-mapping of periodically domain-inverted LiNbO₃ with coherent X-rays

Z. W. Hu, P. A. Thomas, A. Snigirev, I. Snigireva, A. Souvorov, P. G. R. Smith, G. W. Ross & S. Teat

Nature **392**, 690–693 (1998)

The labels for Fig. 2b and c should be interchanged. □